



TUGAS AKHIR - KS 141501

# **ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION (LDA)***

## ***TOPIC ANALYSIS OF PUBLIC INFORMATION IN SOCIAL MEDIA IN SURABAYA BASED ON LATENT DIRICHLET ALLOCATION (LDA) TOPIC MODELLING***

I MADE KUSNANTA BRAMANTYA PUTRA  
NRP 5213 100 035

Dosen Pembimbing :  
Renny Pradina Kusumawardani, S.T., M.T.

JURUSAN SISTEM INFORMASI  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember  
Surabaya 2017

*Halaman ini sengaja dikosongkan*



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TUGAS AKHIR - KS 141501**

# **ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION (LDA)***

**I MADE KUSNANTA BRAMANTYA PUTRA**  
**NRP 5213 100 035**

**Dosen Pembimbing :**  
**Renny Pradina Kusumawardani, S.T., M.T.**

**JURUSAN SISTEM INFORMASI**  
**Fakultas Teknologi Informasi**  
**Institut Teknologi Sepuluh Nopember**  
**Surabaya 2017**

*Halaman ini sengaja dikosongkan*



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**FINAL PROJECT - KS 141501**

# ***TOPIC ANALYSIS OF PUBLIC INFORMATION IN SOCIAL MEDIA IN SURABAYA BASED ON LATENT DIRICHLET ALLOCATION (LDA) TOPIC MODELLING***

**I MADE KUSNANTA BRAMANTYA PUTRA**  
**NRP 5213 100 035**

**Supervisor :**  
**Renny Pradina Kusumawardani, S.T., M.T.**

**JURUSAN SISTEM INFORMASI**  
**Fakultas Teknologi Informasi**  
**Institut Teknologi Sepuluh Nopember**  
**Surabaya 2017**

*Halaman ini sengaja dikosongkan*

## LEMBAR PENGESAHAN

### ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION* (LDA)

#### TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer  
pada

Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh:

**I MADE KUSNANTA BRAMANTYA PUTRA**  
**NRP. 5213 100 035**

Surabaya, 09 Juli 2017

**KEPALA DEPARTEMEN SISTEM INFORMASI**



**Dr. IEN Aris Tjahyanto, M.Kom**  
**NIP.19650310 199102 1 001**

*Halaman ini sengaja dikosongkan*



## LEMBAR PERSETUJUAN

### ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION* (LDA)

#### TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer pada  
Jurusan Sistem Informasi  
Fakultas Teknologi Informasi  
Institut Teknologi Sepuluh Nopember

Oleh :

**I MADE KUSNANTA BRAMANTYA PUTRA**  
**NRP. 5213 100 035**

Disetujui Tim Penguji : Tanggal Ujian : 05 Juli 2017  
Periode Wisuda : September 2017

**Renny Pradina Kusumawardani, S.T., M.T.** (Pembimbing I)

**Nur Aini R., S.Kom., M.Sc.Eng., Ph.D.**

(Penguji I)

**Irmasari Hafidz, S.Kom., M.Sc.**

(Penguji II)

*Halaman ini sengaja dikosongkan*

# **ANALISIS TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI SURABAYA MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION (LDA)***

**Nama** : I Made Kusnanta Bramantya Putra  
**NRP** : 5213 100 035  
**Jurusan** : Sistem Informasi FTIf ITS  
**Pembimbing** : Renny Pradina Kusumawardani, S.T.,  
M.T.  
**Lab** : Akuisisi Data dan Diseminasi Informasi

## **ABSTRAK**

*Radio Suara Surabaya sebagai salah satu radio di Kota Surabaya merupakan radio yang menerapkan format "radio News" dan informasi, dimana informasi yang sering di sampaikan antara lain kondisi lalu lintas, keamanan, dan seputar Kota Surabaya. Radio Suara Surabaya mengembangkan siaran interaktif berbasis jurnalistik masyarakat, dimana agasan ini melibatkan partisipasi warga dalam melaporkan peristiwa kepada penyiar radio yang sedang bertugas. Laporan masyarakat yang masuk kemudian disebarluaskan kembali baik oleh penyiar melalui siaran radio maupun oleh gatekeeper melalui media sosial yang dimiliki Radio Suara Surabaya, baik twitter dan Facebook untuk memberikan informasi kepada masyarakat.*

*Tingginya jumlah laporan perhari yang masuk melalui media sosial dan beragamnya topik dari laporan tersebut menimbulkan kesulitan dalam mengidentifikasi suatu topik dari kumpulan laporan media sosial masyarakat dan menghabiskan banyak waktu jika dilakukan secara manual oleh manusia. Padahal, kumpulan laporan tersebut merupakan sumber data yang sangat berpotensi untuk memberikan informasi apa yang terjadi di Kota Surabaya.*

*Dengan kondisi demikian, dibutuhkan suatu pemodelan topik yang mampu secara otomatis mengklasifikasikan pesan*

*media sosial ke dalam topik-topik yang muncul dari hasil pemodelan. Pemodelan topik dilakukan dengan metode Latent Dirichlet Allocation (LDA), sebuah metode text mining untuk menemukan pola tertentu pada sebuah dokumen dengan menghasilkan beberapa macam topik yang berbeda. Eksperimen pemodelan topik dengan metode LDA menyimpulkan bahwa jumlah topik yang terdapat dalam pesan media sosial adalah 4 topik. Hasil eksperimen ini telah diuji secara mesin dengan nilai perplexity terbaik sebesar 213.41 dan diuji kemudahannya untuk diinterpretasi oleh manusia melalui uji koherensi topik yang terdiri dari Word Intrusion task dan Topic Intrusion Task. Kesimpulan dari uji koherensi topik menyatakan bahwa model yang dihasilkan dengan metode LDA pada studi kasus ini dapat diinterpretasi manusia dengan baik.*

**Kata Kunci:** *Radio Suara Surabaya, pesan media sosial, Latent Dirichlet Allocation (LDA), uji koherensi topik, Word Intrusion Task, Topic Intrusion Task*

# **TOPIK ANALYSIS OF PUBLIK INFORMATION IN SOCIAL MEDIA IN SURABAYA BASED ON LATENT DIRICHLET ALLOCATION (LDA) TOPIC MODELING**

**Student Name : I Made Kusnanta Bramantya Putra**  
**NRP : 5213 100 035**  
**Department : Sistem Informasi FTIf ITS**  
**Supervisor : Renny Pradina Kusumawardani, S.T.,  
M.T.**  
**Lab : Akuisisi Data dan Diseminasi Informasi**

## **ABSTRACT**

*Radio Suara Surabaya as a radio in Surabaya is a radio that applies the format of "Radio News" and information, where information is often conveyed such as traffic conditions, security, and things related to Surabaya. Radio Suara Surabaya is developing an interactive broadcast based on community journalism, where it involves citizen participation in reporting events to radio broadcasters on duty. Incoming community reports are then disseminated both by broadcasters via radio broadcasts and by gatekeepers through social media owned by Radio Suara Surabaya, both twitter and Facebook to provide information to the public.*

*The high number of reports per day coming through social media and the diverse topics of the report make it difficult to identify a topic from a collection of social media reports and spend a lot of time manually done by humans. In fact, the collection of reports is a source of data that has the potential to provide information what is happening in the city of Surabaya.*

*Under such conditions, a topic modeling is required that automatically classifies social media messages into topics arising from the modeling results. Topic modeling is done by the Latent Dirichlet Allocation (LDA) method, a text mining method to find a*

*particular pattern on a document by generating several different topics.*

*Topic modeling experiment with LDA method concludes that the number of topics contained in the message of social media is four topics. The results of this experiment have been tested mechanically with the best perplexity value of 213.41 and tested the ease to be interpreted by humans through the topic coherence test consisting of Word Intrusion task and Topic Intrusion Task. The conclusion of the topic coherence test states that the model produced by the LDA method in this case study can be interpreted human well.*

**Keywords:** *Radio Suara Surabaya, social media messages, Latent Dirichlet Allocation (LDA), topic coherence test, Word Intrusion Task, Topic Intrusion Task*

## KATA PENGANTAR

Puji syukur atas asung kertha wara nugraha Ida Sang Hyang Widhi Wasa yang telah dilimpahkan selama ini sehingga penulis mendapatkan kelancaran dalam menyelesaikan tugas akhir dengan judul :

### **Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan *Latent Dirichlet Allocation (LDA)***

sebagai salah satu syarat kelulusan pada Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember Surabaya.

Penyusunan tugas akhir ini senantiasa mendapatkan dukungan dari berbagai pihak baik dalam bentuk doa, motivasi, semangat, kritik, saran dan berbagai bantuan lainnya. Untuk itu, secara khusus penulis akan menyampaikan ucapan terima kasih yang sedalam-dalamnya kepada:

- 1) Segenap keluarga besar terutama kedua orang tua dan kakak penulis, Bapak Putu Arya Mertasana, Ibu Made Mahayuni, dan Ni Luh Putu Nurindah Sukmawati yang senantiasa mendoakan, memberikan motivasi dan semangat, sehingga penulis mampu menyelesaikan pendidikan S1 ini dengan baik.
- 2) Bapak Dr. Ir. Aris Tjahyanto, M.Kom., selaku Ketua Jurusan Sistem Informasi ITS, Bapak Nisfu Asrul Sani, S.Kom, M.Sc selaku KaProdi S1 Sistem Informasi ITS selama penulis menjalani kuliah, Seluruh dosen pengajar beserta staf dan karyawan di Jurusan Sistem Informasi, FTIF ITS Surabaya yang telah menjalankan perannya sebaik mungkin sehingga penulis dapat mengenyam pendidikan S1 dengan baik.
- 3) Renny Pradina Kusumawardani, S.T., M.T selaku dosen pembimbing yang telah banyak meluangkan waktu untuk membimbing, mengarahkan, dan mendukung dengan memberikan ilmu, petunjuk, dan motivasi dalam penyelesaian Tugas Akhir

- 4) Bapak Prof. Ir. Arif Djunaidy M.Sc., Ph.D. sebagai dosen wali penulis selama menempuh pendidikan di Jurusan Sistem Informasi.
- 5) Ibu Nur Aini Rakhmawati, S.Kom., M.Sc.Eng., Ph.D., dan Ibu Irmasari Hafidz, S.Kom., M.Sc. selaku dosen penguji yang telah memberikan kritik, saran, dan masukan yang dapat menyempurnakan Tugas Akhir ini.
- 6) Kawan-kawan seperjuangan mahasiswa Jurusan Sistem Informasi, terutama angkatan 2013 (Beltranis) serta kawan-kawan mahasiswa Hindu yang tergabung dalam Tim Pembina Kerohanian Hindu ITS terutama angkatan 2013
- 7) Rekan-rekan HMSI Muda Berkarya dan HMSI Revolution khususnya departemen PSDM Teladan yang telah menjadi naungan menempa diri dalam kebersamaan dalam belajar dan berkarya.
- 8) Rekan-rekan pemandu Ekspresi, “Sahabat Sambat” atas semangat dan kebersamaan dalam masa-masa penyusunan tugas akhir.
- 9) Kawan-kawan “Kontrakan Uwus” sebagai lingkungan terdekat yang telah kebersamai selama di Surabaya
- 10) Rekan-rekan Statistika ITS, Endah Setyowati, Vida Faiza Rochmah, dan Desak Pratiwi yang telah banyak membantu pengerjaan tugas akhir ini, khususnya pada bagian yang melibatkan ilmu statistika
- 11) Rekan penulis, Ni Ketut Ratih Larasati yang selalu memotivasi dan memberikan bantuan serta dukungan dan juga menghibur penulis selama pengerjaan tugas akhir ini.
- 12) Serta semua pihak yang telah membantu dalam pengerjaan Tugas Akhir ini yang belum mampu penulis sebutkan diatas.

Terima kasih atas segala bantuan, dukungan, serta doanya. Semoga Ida Sang Hyang Widhi Wasa, Tuhan Yang Maha Esa



senantiasa melimpahkan anugerah serta membalas kebaikan yang telah diberikan kepada penulis.

Penulis menyadari bahwa masih terdapat kekurangan dalam penyusunan tugas akhir ini, oleh karena itu penulis mengharapkan saran dan kritik yang membangun demi kebaikan penulis dan tugas akhir ini

Akhir kata, penulis berharap bahwa tugas akhir ini dapat memberikan kebermanfaatan bagi pembacanya

Surabaya, Juli 2017

Penulis

*Halaman ini sengaja dikosongkan*

## DAFTAR ISI

ABSTRAK .....	xi
ABSTRACT .....	xiii
KATA PENGANTAR .....	xv
DAFTAR ISI .....	xix
DAFTAR GAMBAR .....	xxiii
DAFTAR KODE .....	xxv
DAFTAR TABEL .....	xxvii
DAFTAR DIAGRAM .....	xxxi
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Perumusan Masalah .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	4
1.6. Relevansi .....	4
BAB II TINJAUAN PUSTAKA .....	5
2.1. Studi Sebelumnya .....	5
2.2. Dasar Teori .....	9
2.2.1. <i>Topic modeling</i> .....	9
2.2.2. <i>Latent Dirichlet Allocation</i> .....	11
2.2.3. Validasi Topik dengan <i>Perplexity</i> dan <i>Topic coherence</i> .....	13
2.2.4. Radio Suara Surabaya FM .....	17
BAB III METODOLOGI PENELITIAN .....	19
3.1. Identifikasi Masalah .....	20
3.2. Studi Literatur .....	20
3.3. Mempersiapkan Data .....	21
3.4. <i>Topic modeling</i> dengan <i>Latent Dirichlet Allocation</i> .....	22
3.4.1. Pra-pemrosesan <i>Corpus</i> .....	23
3.4.2. Membentuk Model Topik .....	29
3.4.3. Validasi Model Topik .....	31
3.5. Uji Koherensi Topik .....	31
3.6. Menganalisis Topik Model .....	32
BAB IV PERANCANGAN .....	33
4.1. Pengambilan Data .....	33

4.2.	Observasi Data Mentah .....	36
4.3.	Metodologi Implementasi Penelitian.....	38
4.3.1.	Mempersiapkan Data .....	38
4.3.2.	Pra-Proses Data .....	39
4.3.3.	<i>Topic modeling</i> .....	41
4.3.4.	Validasi Topik Model.....	43
4.3.5.	Uji Koherensi Topik .....	43
BAB V	IMPLEMENTASI .....	49
5.1.	Perangkat Penelitian.....	49
5.2.	Mempersiapkan Data .....	50
5.2.1.	Memuat Data .....	50
5.2.2.	Membersihkan Data.....	51
5.3.	Pra-Proses Data .....	51
5.3.1.	Pendefinisian <i>Stopwords</i> .....	52
5.3.2.	Menghapus <i>Stopwords</i> dari Dokumen .....	52
5.3.3.	Mengkonversi Dokumen ke Dalam bentuk <i>Dictionary</i> .....	53
5.3.4.	Mengkonversi <i>Dictionary</i> dan <i>List</i> ke dalam Matriks Dokumen.....	54
5.4.	Pemodelan Topik dengan <i>Latent Dirichlet Allocation</i> .....	54
5.4.1.	Alur Pemodelan Topik dengan <i>Latent Dirichlet Allocation</i> .....	55
5.4.2.	Eksperimen Pemodelan Topik dengan <i>LDA</i> ...	57
5.5.	Validasi Model Topik .....	59
5.5.1.	Klasifikasi Dokumen ke dalam Topik dengan Metode <i>LDA</i> .....	59
5.5.2.	Analisis Distribusi Probabilitas Dokumen Per Topik .....	62
5.5.3.	Visualisasi Distribusi Probabilitas Dominan dari Dokumen per Topik.....	66
5.6.	Uji Koherensi Topik.....	68
5.6.1.	Menampilkan Topik .....	68
5.6.2.	Penyusunan Materi Kuesioner .....	69
5.6.3.	Sistematika Kuesioner Uji Koherensi Topik.	107
BAB VI	HASIL DAN PEMBAHASAN .....	111
6.1.	Mempersiapkan Data .....	111

6.1.1.	<i>Loading Data</i> .....	111
6.1.2.	Pembersihan Data.....	111
6.2.	Pra-Proses Data.....	111
6.2.1.	Pendefinisian <i>Stopwords</i> .....	111
6.3.	Pembentukan Model <i>LDA</i> .....	114
6.3.1.	Hasil Pembentukan Model <i>LDA</i> dengan <i>Stemming</i> .....	114
6.3.2.	Hasil Pembentukan Model <i>LDA</i> tanpa <i>Stemming</i> .....	115
6.4.	Validasi Model Topik .....	116
6.4.1.	Penentuan Jumlah Iterasi .....	116
6.4.2.	Penentuan Jumlah Topik .....	117
6.4.3.	Jumlah Distribusi Dokumen Per Topik.....	120
6.4.4.	Visualisasi Distribusi Probabilitas Dokumen per Topik.....	122
6.5.	Hasil Uji Koherensi Topik .....	131
6.5.1.	Histogram.....	131
6.5.2.	Uji Hipotesis.....	136
6.5.3.	Uji ANOVA ( <i>Analysis of Variance</i> ) .....	150
BAB VII KESIMPULAN DAN SARAN.....		159
7.1.	Kesimpulan.....	159
7.2.	Saran.....	160
DAFTAR PUSTAKA .....		163
BIODATA PENULIS .....		167
LAMPIRAN A.....		1
LAMPIRAN B .....		1

*Halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

Gambar 2.1 Konsep <i>Topic modeling</i> menurut Blei [13].....	10
Gambar 2.2 Visualisasi <i>Topic modeling</i> dengan Metode <i>LDA</i> [12].....	12
Gambar 2.3 Ilustrasi <i>Word Intrusion task</i> dan <i>Topic Intrusion Task</i> [13] .....	15
Gambar 3.1 Hasil <i>Crawling</i> Akun <i>Twitter @e100ss</i> .....	21
Gambar 3.2 Hasil <i>Crawling</i> Akun <i>Facebook E100</i> .....	22
Gambar 4.1 Contoh data pada tabel <i>fb_test5</i> .....	34
Gambar 4.2 Contoh data pada tabel <i>fb_comments</i> .....	35
Gambar 4.3 Contoh data pada tabel <i>tw_test2</i> .....	36
Gambar 5.1 Bagian judul kuesioner uji koherensi topik.....	107
Gambar 5.2 Bagian deskripsi pembuka kuesioner uji koherensi topik bagian <i>Word Intrusion Task</i> .....	108
Gambar 5.3 Bagian pertanyaan kuesioner uji koherensi topik bagian <i>Word Intrusion Task</i> .....	108
Gambar 5.4 Bagian deskripsi pembuka kuesioner uji koherensi topik bagian <i>Topic Intrusion Task</i> .....	109
Gambar 5.5 Bagian pertanyaan kuesioner uji koherensi topik bagian <i>Topic Intrusion task</i> Eksperimen 1 .....	110
Gambar 5.6 Bagian pertanyaan kuesioner uji koherensi topik bagian <i>Topic Intrusion task</i> Eksperimen 2 .....	110
Gambar 6.1 Uji <i>variance</i> WIT dan TIT 1 .....	138
Gambar 6.2 Uji <i>means</i> WIT dan TIT 1.....	139
Gambar 6.3 Uji <i>variance</i> TIT 1 dan TIT 2 .....	140
Gambar 6.4 Uji <i>Means</i> TIT 1 dan TIT 2 .....	141
Gambar 6.5 Uji <i>variance</i> WIT dan TIT 2 .....	142
Gambar 6.6 Uji <i>Means</i> WIT 1 dan TIT 2 .....	143
Gambar 6.7 Uji <i>variance</i> WIT - <i>Stem</i> dan WIT + <i>Stem</i> .....	144
Gambar 6.8 Uji <i>Means</i> WIT - <i>Stem</i> dan WIT + <i>Stem</i> .....	145
Gambar 6.9 Uji <i>variance</i> TIT 1 - <i>Stem</i> dan TIT 1 + <i>Stem</i> .....	146
Gambar 6.10 Uji <i>Means</i> TIT 1 - <i>Stem</i> dan TIT 1 + <i>Stem</i> .....	147
Gambar 6.11 Uji <i>variance</i> TIT 2 - <i>Stem</i> dan TIT 2 + <i>Stem</i> ...	148
Gambar 6.12 Uji <i>Means</i> TIT 2 - <i>Stem</i> dan TIT 2 + <i>Stem</i> .....	149
Gambar 6.13 Informasi pengelompokan sampel pada <i>Word Intrusion Task</i> .....	152

Gambar 6.14 Perbandingan berpasangan setiap sampel untuk <i>Word Intrusion Task</i> .....	153
Gambar 6.15 Informasi pengelompokan sampel pada <i>Topic Intrusion task 1</i> .....	154
Gambar 6.16 Perbandingan berpasangan setiap sampel untuk <i>Topic Intrusion task 1</i> .....	155
Gambar 6.17 Informasi pengelompokan sampel pada <i>Topic Intrusion task 2</i> .....	156
Gambar 6.18 Perbandingan berpasangan setiap sampel untuk <i>Topic Intrusion task 2</i> .....	157



## DAFTAR KODE

Kode 5.1 Memuat ( <i>Loading</i> ) Data.....	50
Kode 5.2 Membersihkan data.....	51
Kode 5.3 Pendefinisian <i>stopwords</i> .....	52
Kode 5.4 Menghapus <i>stopwords</i> dari dokumen.....	53
Kode 5.5 Mengkonversi dokumen ke dalam bentuk <i>dictionary</i> .....	54
Kode 5.6 Mengkonversi <i>dictionary</i> dan <i>list</i> ke dalam matriks dokumen.....	54
Kode 5.7 <i>Loading Dictionary</i> dan <i>Corpus</i> .....	55
Kode 5.8 Eksperimen pada <i>input parameter LDA</i> .....	56
Kode 5.9 Pendokumentasian <i>Logging</i> .....	56
Kode 5.10 Penentuan jumlah iterasi .....	58
Kode 5.11 Penentuan jumlah topik.....	59
Kode 5.12 Menyimpan model .....	59
Kode 5.13 Import <i>library</i> .....	60
Kode 5.14 <i>Loading input</i> .....	60
Kode 5.15 Operasi data .....	62
Kode 5.16 menampilkan distribusi dokumen dalam topik.....	62
Kode 5.17 Import <i>library</i> .....	63
Kode 5.18 <i>Loading input</i> .....	63
Kode 5.19 Operasi data .....	65
Kode 5.20 Menampilkan jumlah distribusi dokumen per topik .....	66
Kode 5.21 <i>Source code</i> distribusi probabilitas dominan dari dokumen per topik dengan metode <i>LDA</i> .....	67
Kode 5.22 Histogram distribusi probabilitas dominan dari dokumen per topik dengan metode <i>LDA</i> .....	67
Kode 5.23 Menampilkan topik .....	68

*Halaman ini sengaja dikosongkan*

## DAFTAR TABEL

Tabel 2.1 Penelitian Sebelumnya .....	5
Tabel 3.1 Contoh Data Sebelum dan Sesudah Melalui Tahap <i>Lowercase</i> .....	24
Tabel 3.2 Contoh Data Sebelum dan Sesudah Melalui Tahap Tokenization .....	25
Tabel 3.3 Contoh Kelas Kata .....	26
Tabel 3.4 Contoh daftar <i>stopwords</i> berdasarkan penelitian Fadillah Z Tala [23] .....	27
Tabel 3.5 Contoh Data Sebelum dan Sesudah Melalui Tahap Penghapusan <i>Stopwords</i> .....	27
Tabel 3.6 Contoh Data Sebelum dan Sesudah Melalui Tahap <i>Stemming</i> .....	28
Tabel 3.7 Contoh Luaran Model Topik Hasil Eksperimen ....	29
Tabel 4.1 Tabel Sumber Data Masukan .....	33
Tabel 4.2 Tabel Atribut pada Tabel fb_test5 .....	34
Tabel 4.3 Tabel Atribut pada Tabel fb_comments .....	34
Tabel 4.4 Tabel Atribut pada Tabel tw_test2 .....	35
Tabel 4.5 Justifikasi data masukan .....	36
Tabel 5.1 Deret Akumulatif Topik #0 .....	69
Tabel 5.2 Pengacakan dengan <i>Roulette wheel</i> pada Topik #0	70
Tabel 5.3 Deret Akumulatif Topik #1 .....	71
Tabel 5.4 Pengacakan dengan <i>Roulette wheel</i> pada Topik #1	72
Tabel 5.5 Deret Akumulatif Topik #2 .....	72
Tabel 5.6 Pengacakan dengan <i>Roulette wheel</i> pada Topik #2	73
Tabel 5.7 Deret Akumulatif Topik #3 .....	74
Tabel 5.8 Pengacakan dengan <i>Roulette wheel</i> pada Topik #3	75
Tabel 5.9 Deret Akumulatif Topik #0 .....	76
Tabel 5.10 Pengacakan dengan <i>Roulette wheel</i> pada Topik #0 .....	77
Tabel 5.11 Deret Akumulatif Topik #1 .....	78
Tabel 5.12 Pengacakan dengan <i>Roulette wheel</i> pada Topik #1 .....	78
Tabel 5.13 Deret Akumulatif Topik #2 .....	79
Tabel 5.14 Pengacakan dengan <i>Roulette wheel</i> pada Topik #2 .....	80

Tabel 5.15 Deret Akumulatif Topik #3 .....	81
Tabel 5.16 Pengacakan dengan <i>Roulette wheel</i> pada Topik #3 .....	82
Tabel 5.17 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> dengan <i>Stem</i> .....	83
Tabel 5.18 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> dengan <i>Stem</i> Eksperimen 2 untuk Topik #0.....	84
Tabel 5.19 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> dengan <i>Stem</i> Eksperimen 2 untuk Topik #1 .....	85
Tabel 5.20 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> dengan <i>Stem</i> Eksperimen 2 untuk Topik #2.....	86
Tabel 5.21 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> dengan <i>Stem</i> Eksperimen 2 untuk Topik #3.....	87
Tabel 5.22 Daftar Dokumen <i>Topic Intrusion task</i> dengan <i>Stem</i> .....	88
Tabel 5.23 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> tanpa <i>Stem</i> Eksperimen 1 .....	93
Tabel 5.24 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> tanpa <i>Stem</i> Eksperimen 2 Topik #0.....	94
Tabel 5.25 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> tanpa <i>Stem</i> Eksperimen 2 Topik #1 .....	95
Tabel 5.26 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> tanpa <i>Stem</i> Eksperimen 2 Topik #2.....	96
Tabel 5.27 Daftar Opsi Kuesioner <i>Topic Intrusion task</i> tanpa <i>Stem</i> Eksperimen 2 Topik #3.....	97
Tabel 5.28 Daftar Dokumen <i>Topic Intrusion task</i> tanpa <i>Stem</i>	98
Tabel 6.1 Data mentah yang dimuat sebagai sumber data masukan .....	111
Tabel 6.2 Hasil pembersihan data.....	111
Tabel 6.3 Kelas kata yang memenuhi syarat sebagai <i>stopword</i> .....	112
Tabel 6.4 Hasil pembersihan data melalui <i>stopwords</i> .....	113
Tabel 6.5 Hasil Pembentukan Model <i>LDA</i> dengan <i>Stemming</i> .....	114
Tabel 6.6 Hasil Pembentukan Model <i>LDA</i> tanpa <i>Stemming</i>	115
Tabel 6.7 Distribusi Dokumen Per Topik dengan <i>Stemming</i>	121
Tabel 6.8 Distribusi Dokumen Per Topik tanpa <i>Stemming</i> ..	121

Tabel 6.9 Tabel Pendahuluan Uji Hipotesis .....	137
Tabel 6.10 Tabel Rekapitulasi Uji <i>Variance</i> .....	149
Tabel 6.11 Tabel Rekapitulasi Uji <i>Means</i> .....	150
Tabel 6.12 Tabel Hasil Uji <i>Variance</i> .....	151
Tabel 0.1 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik dengan Stemming.....	19
Tabel 0.2 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik dengan Stemming (lanjutan) .....	20
Tabel 0.3 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik tanpa Stemming .....	21
Tabel 0.4 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik tanpa Stemming (lanjutan) .....	22
Tabel 0.5 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden Word Intrusion Task .....	28
Tabel 0.6 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden Topic Intrusion Task 1 .....	31
Tabel 0.7 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden Topic Intrusion Task 2 .....	35

*Halaman ini sengaja dikosongkan*

## DAFTAR DIAGRAM

Diagram 3.1 Metodologi Penelitian.....	19
Diagram 3.2 Tahap <i>Topic modeling</i> dengan <i>Latent Dirichlet Allocation (LDA)</i> .....	23
Diagram 3.3 Sub-aktivitas dari tahap pra-pemrosesan <i>corpus</i> .....	24
Diagram 4.1 Alur mempersiapkan data .....	39
Diagram 4.2 Alur praproses data .....	40
Diagram 4.3 Alur pembentukan Dictionary dan Corpus .....	41
Diagram 4.4 Alur topic modeling dengan LDA .....	42
Diagram 4.5 Penjabaran bagian-bagian kuesioner uji koherensi topik .....	43
Diagram 6.1 Analisis Nilai <i>Perplexity</i> untuk Penentuan Jumlah Iterasi .....	116
Diagram 6.2 Rata-rata nilai <i>Perplexity</i> 30 percobaan .....	117
Diagram 6.3 Standar Deviasi Nilai <i>Perplexity</i> 30 Percobaan .....	118
Diagram 6.4 Rata-rata nilai <i>Perplexity</i> 30 percobaan .....	119
Diagram 6.5 Standar Deviasi Nilai <i>Perplexity</i> 30 Percobaan .....	120
Diagram 6.6 Histogram distribusi probabilitas dokumen pada Topik #0 .....	123
Diagram 6.7 Histogram distribusi probabilitas dokumen pada Topik #1 .....	124
Diagram 6.8 Histogram distribusi probabilitas dokumen pada Topik #2 .....	125
Diagram 6.9 Histogram distribusi probabilitas dokumen pada Topik #3 .....	126
Diagram 6.10 Histogram distribusi probabilitas dokumen pada Topik #0 .....	127
Diagram 6.11 Histogram distribusi probabilitas dokumen pada Topik #1 .....	128
Diagram 6.12 Histogram distribusi probabilitas dokumen pada Topik #2 .....	129

Diagram 6.13 Histogram distribusi probabilitas dokumen pada Topik #3 .....	130
Diagram 6.14 Histogram skor <i>Word Intrusion task</i> berbasis pertanyaan .....	132
Diagram 6.15 Histogram skor <i>Topic Intrusion task</i> 1 berbasis pertanyaan .....	132
Diagram 6.16 Histogram skor <i>Topic Intrusion task</i> 2 berbasis pertanyaan .....	133
Diagram 6.17 Histogram skor <i>Word Intrusion task</i> berbasis responden .....	134
Diagram 6.18 Histogram skor <i>Topic Intrusion task</i> 1 berbasis responden .....	135
Diagram 6.19 Histogram skor <i>Topic Intrusion task</i> 2 berbasis responden .....	136



# **BAB I**

## **PENDAHULUAN**

Pada bab pendahuluan akan diuraikan proses identifikasi masalah penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, manfaat kegiatan tugas akhir dan relevansi terhadap pengerjaan tugas akhir. Berdasarkan uraian pada bab ini, harapannya gambaran umum permasalahan dan pemecahan masalah pada tugas akhir dapat dipahami.

### **1.1. Latar Belakang**

Kota Surabaya merupakan ibu kota Provinsi Jawa Timur Indonesia, sekaligus menjadi kota metropolitan terbesar di provinsi tersebut. Surabaya sebagai kota terbesar kedua di Indonesia setelah Jakarta juga merupakan pusat bisnis, perdagangan, industri, dan pendidikan di Jawa Timur serta wilayah Indonesia lainnya di bagian timur. Secara geografis, Kota Surabaya terletak di 07° 15' 50" Lintang Selatan dan 112° 45' 03" Bujur Timur, atau di tepi pantai utara Pulau Jawa dan berhadapan dengan Selat Madura serta Laut Jawa. Surabaya memiliki luas sekitar 333,063 km<sup>2</sup> dengan penduduknya berjumlah 2.909.257 jiwa (2015) [1].

Salah satu keunikan Kota Surabaya adalah adanya Radio Suara Surabaya FM (SSFM) yang memiliki slogan *News, Interaktif dan Solutif*, sehingga dalam operasionalnya, tidak menyiarkan musik dan hiburan seperti radio lainnya, namun SSFM ini berhasil menjadi penghubung masyarakat yang memonitor kondisi keseharian warga Kota Surabaya dan sekitarnya. SSFM mampu meningkatkan partisipasi pendengarnya sebagai “reporter jalanan” dalam bentuk siaran interaktif dengan penyiar dan *gatekeeper*. Konsistensi SSFM sebagai radio berita interaktif telah mampu menjadi lembaga yang mewadahi interaksi sosial sehari-hari dalam masyarakat. [2]

Tidak hanya melalui radio, Radio Suara Surabaya FM juga mengikuti perkembangan jaman di masyarakat dengan

merambah media sosial. Akun e100ss adalah akun *Twitter* dan *Facebook* resmi Radio Suara Surabaya yang dimanfaatkan sebagai salah satu saluran media komunikasi untuk berbagi berita secara *real-time* dengan pendengar untuk mendukung siaran Suara Surabaya FM (SSFm) [3]

Jumlah informasi publik media sosial yang di-*post* melalui media sosial resmi Radio Suara Surabaya FM sangat tinggi, hal ini ditunjukkan dengan jumlah rata-rata *tweet* dalam sehari dari akun @e100ss sejumlah 192,2<sup>1</sup>, sementara jumlah *post* dalam sehari dari akun E100 sejumlah 126,18 *post*<sup>2</sup>, yang mana kedua data tersebut belum termasuk pesan dari pendengar atau pengguna yang ditujukan kepada media sosial resmi Radio Suara Surabaya FM. Dengan jumlah *post* yang tinggi, terdapat pula hal lain yang perlu disoroti, yaitu pola pesan yang tidak sesuai dengan ejaan yang disempurnakan dan topik yang beragam dan sangat cepat berubah dari satu waktu ke waktu, sehingga untuk mengidentifikasi suatu topik dari kumpulan *tweet* tersebut sangat sulit dan menghabiskan banyak waktu jika dilakukan secara manual oleh manusia. Padahal, kumpulan *post* tersebut merupakan sumber data yang sangat berpotensi untuk memberikan informasi apa yang terjadi di Kota Surabaya.

Berdasarkan permasalahan tersebut, penelitian ini akan menawarkan solusi dalam melakukan analisis *Topic modeling* terhadap pesan media sosial yang disampaikan melalui akun e100ss. Analisis *Topic modeling* digunakan untuk mengetahui topik-topik apa saja yang sering muncul, sehingga memudahkan dalam mengetahui apa yang sedang terjadi di kota Surabaya. Penelitian ini menggunakan pembelajaran *machine learning* dengan metode *Latent Dirichlet Allocation (LDA)* [4]. Metode *LDA* merupakan sebuah metode *text mining* untuk menemukan pola tertentu pada sebuah dokumen dengan menghasilkan beberapa macam topik yang berbeda, sehingga

---

<sup>1</sup> Menurut data pada bulan Desember 2016 berdasarkan website analisis *Twitter*, *tweetchup.com*

<sup>2</sup> Menurut data pada bulan Desember 2016 berdasarkan website analisis *Facebook*, *likealyzer.com*

tidak secara spesifik mengelompokkan dokumen kedalam sebuah topik tertentu [4]. Topik yang muncul dari pengolahan data tersebut selanjutnya akan dilakukan uji koherensi topik, yaitu uji keterkaitan dari uraian probabilitas kata-kata yang ditemukan satu sama lain dalam menyusun suatu topik [5]. Hasil dari analisis *Topic modeling* menggunakan *LDA* diharapkan dapat membantu dalam memahami apa yang sedang terjadi di Surabaya secara lebih ringkas.

## 1.2. Perumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan permasalahan yang menjadi fokus dan akan diselesaikan dalam tugas akhir ini antara lain:

1. Bagaimana melakukan *Topic modeling* untuk menganalisis topik-topik apa yang sedang dibahas pada media sosial informasi publik di Surabaya?
2. Bagaimana mengukur tingkat koherensi topik yang menjadi luaran *Topic modeling*?

## 1.3. Batasan Masalah

Dari permasalahan yang disebutkan di atas, batasan masalah dalam tugas akhir ini adalah:

1. Studi kasus yang digunakan berfokus pada lingkup kota Surabaya
2. Data bersumber dari jejaring sosial resmi Radio Suara Surabaya FM, yaitu akun @e100ss untuk jejaring sosial *Twitter* dan akun E100 untuk jejaring sosial *Facebook* dari bulan Agustus 2015 sampai Desember 2016
3. Jenis data yang dianalisis menggunakan metode *LDA* berupa data teks

## 1.4. Tujuan Penelitian

Tujuan dalam tugas akhir ini adalah:

1. Melakukan analisis *Topic modeling* untuk mengetahui topik-topik apa yang sedang

dibahas pada media sosial informasi publik di Surabaya.

2. Melakukan validasi pada luaran dari *Topic modeling* untuk mengukur kinerja metode *LDA* pada dokumen.

### 1.5. Manfaat Penelitian

Manfaat yang diharapkan dapat diperoleh dari tugas akhir ini adalah:

#### Bagi Penulis:

- Memahami metode *Topic modeling* dengan algoritma *Latent Dirichlet Allocation (LDA)* dan mampu menerapkannya pada studi kasus pesan informasi publik media sosial
- Melakukan eksperimen untuk mengetahui kinerja pemodelan *Latent Dirichlet Allocation (LDA)* dengan karakteristik data pesan media sosial

### 1.6. Relevansi

Topik pada tugas akhir ini merupakan *Topic modeling*, sehingga masih berkaitan dengan topik yang berada pada Laboratorium Akuisisi Data dan Diseminasi Informasi. Adapun mata kuliah yang relevan dengan tugas akhir ini yaitu Algoritma dan Pemrograman, Matematika Diskrit, Sistem Cerdas, Penggalan Data dan Analitika Bisnis dan Statistika.

## BAB II TINJAUAN PUSTAKA

Pada bagian tinjauan pustaka ini, akan dijelaskan mengenai referensi-referensi yang terkait dalam penyusunan tugas akhir ini.

### 2.1. Studi Sebelumnya

Pada pengerjaan tugas akhir ini terdapat beberapa penelitian sebelumnya yang dijadikan acuan. Tabel 2.1 menampilkan daftar penelitian sebelumnya yang mendasari tugas akhir ini:

**Tabel 2.1 Penelitian Sebelumnya**

Judul penelitian	Metode	Penulis	Hasil yang Didapatkan
Identifikasi Topik Informasi Publik Media Sosial di Kota Surabaya Berdasarkan Klasterisasi Teks pada <i>Twitter</i> dengan	Klasterisasi dengan Algoritma <i>K-Means</i>	Moh. Hasan Basri	Penelitian yang dilakukan berhasil mengidentifikasi 31 jenis topik yang dibahas pada informasi publik media sosial Surabaya dalam bentuk hierarki topik. Identifikasi dilakukan dengan menggunakan algoritma <i>SVM "linear"</i> dengan jumlah label 13 memiliki nilai akurasi tertinggi yaitu 96,79% [6]

Menggunakan Algoritma <i>K-Means</i>			
<i>A Text mining Research Based On LDA Topic modeling</i>	<i>Latent Dirichlet Allocation Topic modeling</i>	Zhou Tong dan Haiyi Zhang	Paper ini memperkenalkan <i>Text mining</i> dengan <i>LDA Topic modeling</i> sebagai metodenya, dimana eksperimen dilakukan pada dua tipe dokumen, yaitu artikel <i>Wikipedia</i> dan <i>tweet</i> dari pengguna <i>Twitter</i> . Garis besar penelitian ini membahas tentang gambaran umum <i>text mining</i> dengan metode <i>LDA</i> , <i>pre-processing</i> , <i>model training</i> dan hasil analisis. [7]
<i>Topik Models As A Novel Approach To Identify Themes In Content Analysis: The Example Of Organizational Research Methods</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Ajai Gaur	Penelitian ini mendemonstrasikan <i>LDA Topic modeling</i> sebagai metode baru dalam analisis data dalam bentuk teks. Adapun data teks yang dianalisis adalah 421 artikel yang dipublikasikan pada <i>Organization Research Method (ORM)</i> yang berhasil mengungkapkan 15 topik, yang mana hasil analisis tersebut cukup sesuai dengan hasil kajian oleh manusia. [8]
<i>Automatic</i>	<i>TF-IDF</i> dan	Willyh	Penelitian ini bertujuan untuk menghasilkan sebuah

<i>Summarization from Indonesian Hashtag on Twitter Using TF-IDF and Phrase Reinforcement Algorithm</i>	<i>Phrase Reinforcement Algorithm</i>	Hariardi, Novita Latief, David Febryanto dan Derwin Suhartono	kesimpulan terkait apa yang diperbincangkan saat ini di Indonesia melalui <i>hashtag</i> pengguna <i>Twitter</i> di Indonesia. Metode analisis yang digunakan adalah dengan mengkombinasikan <i>TF-IDF</i> dan <i>Phrase Reinforcement Algorithm</i> . Penelitian ini membuktikan metode yang digunakan dapat bekerja dengan baik, meskipun secara umum kesimpulan-kesimpulan yang dihasilkan masih berkualitas rendah dikarenakan banyaknya <i>noise</i> pada data, hal ini ditunjukkan dengan nilai <i>precision</i> yaitu 0.327 dan <i>ROUGE-1</i> sebesar 0.3087 [9]
<i>On-line Trend Analysis with Topik Models: #Twitter trends detection topik model online</i>	<i>Online Variant of Latent Dirichlet Allocation (LDA) Topic modeling</i>	Jey Han Lau, Nigel Collier, Timothy Baldwin	Penelitian ini mengusulkan pendekatan berbasis topik model <i>on-line</i> untuk menganalisis tren. Secara umum, pada setiap <i>update</i> , metode ini menghitung evolusi topik untuk mendeteksi topik baru yang muncul dalam koleksi dokumen. Metodologi ini dapat menunjukkan kekuatan dari model dalam mendeteksi dokumen-dokumen individual dengan menggambarkan tren saat ini, dan terus bergerak

			dalam mengolah data mentah <i>Twitter</i> untuk mendeteksi tren. Tren yang ditemukan cukup memuaskan dan sesuai dengan yang sedang dibahas pada <i>Twittersphere</i> [10]
<i>Software Framework for Topic modeling with Large Corpora</i>	Penelitian ini menggunakan metode <i>Vector Space Model (VSM)</i> yaitu <i>Latent Semantik Analysis (LSA)</i> dan <i>Latent Dirichlet Allocation (LDA)</i> dengan <i>framework Gensim</i> dan Bahasa <i>Python</i>	Radim Rehurek dan Petr Sojka	<i>Vector Space Model (VSM)</i> adalah paradigma dalam <i>modelling</i> yang terbukti dan ampuh dalam <i>Natural Language Processing</i> , di mana dokumen direpresentasikan sebagai vektor dalam ruang berdimensi tinggi. Metode yang termasuk dalam <i>VSM</i> cukup beragam, sehingga algoritma yang diterapkan juga beragam. Dengan demikian, diperlukan suatu <i>framework</i> untuk memberikan arahan sebagai solusi dari adanya <i>practical gap</i> antara model matematis, algoritma dan <i>Source code</i> . Penelitian ini mengajukan adanya <i>framework</i> yang mencakup aspek <i>Corpus size independence</i> , <i>Intuitive API</i> , <i>Easy deployment</i> , <i>Cover popular algorithms</i> dengan menggunakan bahasa <i>Python</i> [11]

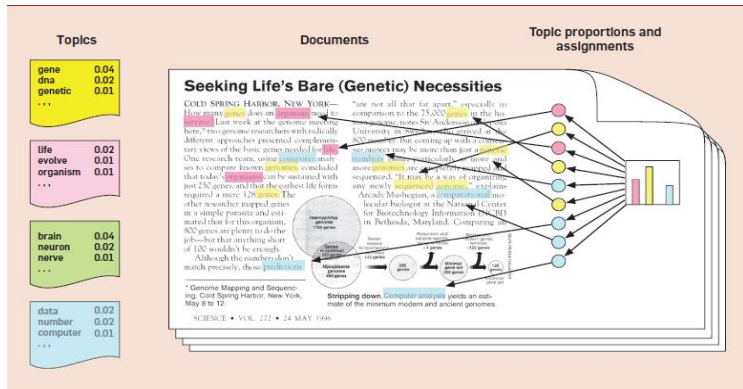


## 2.2. Dasar Teori

### 2.2.1. *Topic modeling*

Konsep *Topic modeling* menurut Blei terdiri dari entitas-entitas yaitu “kata”, “dokumen”, dan “*corpora*”. “Kata” dianggap sebagai unit dasar dari data diskrit dalam dokumen, didefinisikan sebagai item dari kosa kata yang diberi indeks untuk setiap kata unik pada dokumen. “Dokumen” adalah susunan N kata-kata. Sebuah *corpus* adalah kumpulan M dokumen dan *corpora* merupakan bentuk jamak dari *corpus*. Sementara “topik” adalah distribusi dari beberapa kosakata yang bersifat tetap. Secara sederhana, setiap dokumen dalam *corpus* mengandung proporsi tersendiri dari topik-topik yang dibahas sesuai kata-kata yang terkandung di dalamnya [4].

Ide dasar dari *Topic modeling* adalah bahwa sebuah topik terdiri dari kata-kata tertentu yang menyusun topik tersebut, dan dalam satu dokumen memiliki kemungkinan terdiri dari beberapa topik dengan probabilitas masing-masing. Namun secara pemahaman manusia, dokumen-dokumen merupakan objek yang dapat diamati, sedangkan topik, distribusi topik per-dokumen, dan penggolongan setiap kata pada topik per-dokumen merupakan struktur tersembunyi, maka dari itu *Topic modeling* bertujuan untuk menemukan topik dan kata-kata yang terdapat pada topik tersebut [12]. Konsep *Topic modeling* menurut Blei, ditunjukkan pada Gambar 2.1



Gambar 2.1 Konsep *Topic modeling* menurut Blei [13]

Menurut Blei (2012), *Topic modeling* merupakan rangkaian algoritma yang bertujuan untuk menemukan dan memberikan keterangan pada suatu arsip besar dokumen dengan informasi tematik, yaitu pembelajaran terpadu yang menggunakan tema untuk mengaitkan beberapa tema dengan entitas yang dikaitkan. Secara sederhana, *Topic modeling* merupakan algoritma yang bertujuan untuk menemukan topik yang tersembunyi dari rangkaian kata dalam dokumen yang tidak terstruktur. Algoritma *Topic modeling* menganalisis kata-kata dari teks asli untuk menemukan topik yang berada diantara teks tersebut, bagaimana topik topik saling terhubung satu sama lain, dan bagaimana tema-tema tersebut dapat berubah dari waktu ke waktu, sehingga dapat dikembangkan untuk pencarian, ataupun meringkas teks yang terdapat dalam dokumen. [4].

Pendapat lain terkait *Topic modeling* disampaikan oleh Megan R. Brett (2012), yang menyatakan bahwa *Topic modeling* merupakan bentuk *text mining*, sebagai salah satu metode untuk mengidentifikasi pola dalam sebuah *corpus*. *Topic modeling* dapat pula dikatakan sebagai sebuah tool untuk mengubah *corpus* yang berbentuk kumpulan kata, menjadi topik yang dapat menggambarkan *corpus* tersebut [13].

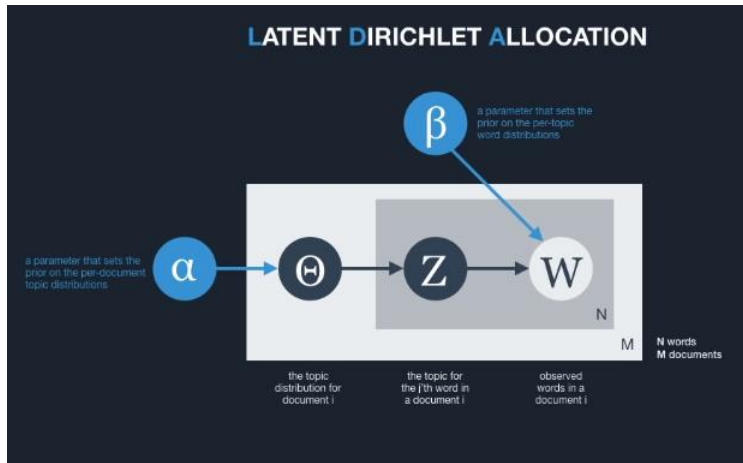
Algoritma pemodelan topik tidak memerlukan penjelasan sebelumnya atau pelabelan karena topik dokumen muncul dari analisis teks asli. Pada umumnya, *Topic modeling* diaplikasikan pada jumlah dokumen yang sangat besar dan dapat mengelola dokumen-dokumen individual sesuai topik yang ditemukan, sehingga *Topic modeling* memungkinkan kita untuk mengatur dan meringkas arsip elektronik pada skala yang tidak mungkin dengan penjelasan manusia secara manual [12].

### **2.2.2. Latent Dirichlet Allocation**

*Latent Dirichlet Allocation (LDA)* merupakan metode Topic modeling dan topik analisis yang paling populer saat ini. *LDA* muncul sebagai salah satu metode yang dipilih dalam melakukan analisis pada dokumen yang berukuran sangat besar. *LDA* dapat digunakan untuk meringkas, melakukan klusterisasi, menghubungkan maupun memproses data yang sangat besar karena *LDA* menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen [14]. Adapun distribusi yang digunakan untuk mendapatkan distribusi topik per-dokumen disebut distribusi Dirichlet, kemudian dalam proses generatif untuk *LDA*, hasil dari Dirichlet digunakan untuk mengalokasikan kata-kata pada dokumen untuk topik yang berbeda. Dalam *LDA*, dokumen-dokumen merupakan objek yang dapat diamati, sedangkan topik, distribusi topik per-dokumen, penggolongan setiap kata pada topik per-dokumen merupakan struktur tersembunyi<sup>3</sup>, maka dari itu, Algoritma ini dinamakan *Latent Dirichlet Allocation (LDA)* [12]. Menurut Blei (2003), *LDA* merupakan model probabilistik generatif dari kumpulan tulisan yang disebut *corpus*. Ide dasar yang diusulkan metode *LDA* adalah setiap dokumen direpresentasikan sebagai campuran acak atas topik yang tersembunyi, yang mana setiap topik memiliki karakter yang ditentukan berdasarkan distribusi kata-kata yang terdapat di dalamnya [4]. Blei merepresentasikan metode *LDA* sebagai model *probabilistic* secara visual seperti pada Gambar 2.2

---

<sup>3</sup> “Latent” merupakan Bahasa Inggris dari kata “Tersembunyi”



**Gambar 2.2 Visualisasi *Topic modeling* dengan Metode LDA [12]**

Sesuai visualisasi model di atas, terdapat tiga tingkatan pada LDA Modeling. Parameter  $\alpha$  dan  $\beta$  merupakan parameter distribusi topik yang berada pada tingkatan *corpus*, yaitu kumpulan dari  $M$  dokumen. Parameter  $\alpha$  digunakan dalam menentukan distribusi topik dalam dokumen, semakin besar nilai  $\alpha$  dalam suatu dokumen, menandakan campuran topik yang dibahas dalam dokumen semakin banyak. Parameter  $\beta$  digunakan untuk menentukan distribusi kata dalam topik. Semakin tinggi nilai  $\beta$ , maka semakin banyak kata-kata yang ada di dalam topik, sedangkan semakin kecil nilai  $\beta$ , maka semakin sedikit kata-kata yang ada di dalam topik sehingga topik tersebut mengandung kata-kata yang lebih spesifik. Variabel  $\theta_m$  adalah variabel yang berada di tingkat dokumen ( $M$ ). Variabel  $\theta$  merepresentasikan distribusi topik untuk dokumen tertentu. Semakin tinggi nilai  $\theta$ , maka semakin banyak topik yang ada di dalam dokumen, sedangkan semakin kecil nilai  $\theta$ , maka dapat dikatakan dokumen tersebut semakin spesifik pada topik tertentu. Variabel  $Z_n$  dan  $W_n$  adalah variabel tingkat kata ( $N$ ). Variabel  $Z$  dan merepresentasikan topik dari kata tertentu pada sebuah dokumen sdangkan variabel  $W$

merepresentasikan kata yang berkaitan dengan topik tertentu yang terdapat dalam dokumen [4]

Secara umum, *LDA* bekerja dengan masukan dokumen-dokumen individual dan beberapa parameter, untuk menghasilkan luaran berupa model yang terdiri dari bobot yang dapat dinormalisasi sesuai probabilitas. Probabilitas ini mengacu pada dua jenis, yaitu jenis (a) probabilitas bahwa suatu dokumen spesifik tertentu menghasilkan topik yang spesifik pula dan jenis (b) probabilitas bahwa topik spesifik tertentu menghasilkan kata-kata spesifik dari sebuah kumpulan kosakata. Probabilitas jenis (a), dokumen yang sudah diberi label dengan daftar topik seringkali dilanjutkan hingga menghasilkan probabilitas jenis (b), yang menghasilkan kata-kata spesifik tertentu. [14]

*LDA* dapat digunakan dalam bidang analisis trend pada media sosial [10], melakukan identifikasi tema pada 421 artikel ilmiah pada *Organizational Research Methods* [8], mendeteksi topik untuk pelacakan konten percakapan [15] dan telah terbukti mampu bekerja dengan baik untuk dokumen panjang seperti artikel *Wikipedia* maupun dokumen pendek seperti *tweet* [7]

### **2.2.3. Validasi Topik dengan *Perplexity* dan *Topic coherence***

*Topic modeling* mempelajari kumpulan dari kata-kata dari sebuah dokumen maupun *corpus* tanpa supervisi manusia (*unsupervised*). Berdasarkan kata-kata yang digunakan yang terdapat dalam dokumen, penggalian relasi topik dilakukan dengan asumsi bahwa satu dokumen mencakup suatu set kecil dari topik yang ringkas, yang mana topik-topik ini perlu dikorelasikan dengan interpretasi manusia [16].

Dalam penelitian ini, akan diterapkan dua metode untuk melakukan validasi topik, yaitu *Perplexity* dan *Topic coherence*.

*Perplexity* menjadi ukuran kualitas standar untuk model topik. *Perplexity* mengukur kemampuan model topik untuk

menggeneralisasi dokumen setelah memperkirakan model menggunakan dokumen pelatihan. *Perplexity* yang lebih rendah berarti kemampuan generalisasi yang lebih baik [27]. Penelitian lain menyebutkan *Perplexity* merupakan ukuran kinerja pemodelan bahasa berdasarkan probabilitas rata-rata yang dikembangkan dalam bidang teori informasi [25]

Untuk menghitung *Perplexity*, diperlukan pemodelan bahasa dan teks yang akan diujikan. *Perplexity* dapat pula digunakan untuk membandingkan berbagai jenis pemodelan Bahasa, namun untuk perbandingan ini, harus menggunakan teks yang sama sebagai teks yang diujikan. Ukuran *vocabulary* dapat dengan mudah terlihat sebagai relevansi dengan *Perplexity* karena dengan berkurangnya kardinalitas dapat secara langsung menurunkan jumlah kemungkinan kata [25].

Chang, dkk mengusulkan bentuk evaluasi topik yang lebih menonjolkan sisi kemudahan dalam interpretasi oleh manusia. Adalah *Topic coherence*, dimana satu set kata-kata yang dihasilkan oleh topik model dinilai berdasarkan tingkat koherensi atau tingkat kemudahannya dalam diinterpretasi manusia [17]. *Topic coherence* mengukur nilai suatu topik dengan mengukur tingkat kesamaan semantik antara kata-kata yang terdapat dalam topik. Pengukuran ini membantu membedakan antara topik yang dapat diinterpretasi secara semantik dengan topik yang memiliki keterkaitan secara statistik [16]. Chang, dkk, dalam penelitiannya tahun 2009 menawarkan task-task untuk mengukur kesuksesan interpretasi topik hasil *Topic modeling*, yaitu *Word Intrusion task* dan *Topic Intrusion Task* [18]. Untuk memudahkan dalam memahami *word intrusion* dan *topic intrusion* ditunjukkan dengan gambar Gambar 2.3

### Word Intrusion

1 / 10  
floppy alphabet computer processor memory disk

2 / 10  
molecule education study university school student

3 / 10  
linguistics actor film comedy director movie

4 / 10  
islands island bird coast portuguese mainland

### Topic Intrusion

6 / 10  
**DOUGLAS HOFSTADTER**  
Douglas Richard Hofstadter (born February 15, 1945 in New York, New York) is an American academic whose research focuses on consciousness, thinking and creativity. He is best known for ", first published in [Show entire excerpt](#)

student	school	study	education	research	university	science	learn
human	life	scientific	science	scientist	experiment	work	idea
play	role	good	actor	star	career	show	performance
write	work	book	publish	life	friend	influence	father

**Gambar 2.3 Ilustrasi Word Intrusion task dan Topic Intrusion Task [13]**

Berdasarkan topik-topik yang sudah dihasilkan melalui *Topic modeling*, *Word Intrusion task* bekerja dengan memilih salah satu dari beberapa topik yang terdapat dalam dokumen dipilih secara acak, kemudian diantara seluruh kata-kata yang terdapat dalam topik disisipkan kata yang memiliki probabilitas rendah pada topik tersebut. Pada akhirnya, pengujian dilakukan dengan cara meminta beberapa responden untuk menebak manakah kata yang tidak termasuk dalam topik tersebut. *Topic Intrusion task* menguji apakah topik yang dihasilkan dari dokumen melalui *Topic modeling* sesuai dengan topik yang dihasilkan dari pandangan manusia terhadap dokumen yang

sama. Pengujian kesesuaian topik dilakukan dengan menyajikan cuplikan dokumen yang disertai beberapa pilihan topik kepada beberapa responden, yang mana satu dari beberapa topik tersebut merupakan topik yang memiliki probabilitas rendah dalam dokumen [18].

### **2.2.3.1. Analisis lanjutan**

Sebagai bentuk tindak lanjut dari metode uji koherensi topik, diperlukan analisis pada berbagai tingkatan, baik tingkatan task, perbedaan perlakuan maupun tingkatan topik. Analisis pada tingkatan task dan perbedaan perlakuan, dilakukan dengan uji hipotesis, dengan dugaan sementara nilai rata-rata untuk sampel yang diujikan dianggap sama. Analisis pada tingkatan topik, dilakukan dengan menguji derajat pemahaman responden untuk setiap topik pada masing-masing task berdasarkan hasil kuesioner. Pengujian derajat pemahaman responden dilakukan dengan menghitung signifikansi nilai dari rataan untuk setiap topik.

Untuk melakukan analisis terhadap signifikansi nilai dari beberapa rataan dilakukan dengan menggunakan uji ANOVA (analysis of varians) satu arah. Terdapat berbagai jenis metode uji perbandingan berganda yang dapat digunakan untuk mengidentifikasi nilai rata-rata yang berbeda dari yang lainnya. Untuk data dengan varians yang sama digunakan metode Tukey [19], sementara untuk data dengan varians yang berbeda digunakan metode Games-Howell [20]. Kedua tes ini membandingkan perbedaan antara masing-masing nilai rataan secara berpasangan, kemudian hasilnya disajikan sebagai matriks yang menunjukkan hasil untuk masing-masing pasangan dalam bentuk p-value dan confidence interval. Uji perbandingan Tukey mengasumsikan bahwa data dari berbagai kelompok objek amatan memiliki distribusi normal dan standar deviasi sama untuk setiap kelompok. Signifikansi nilai ditampilkan dalam kelompok yang dinotasikan dalam bentuk huruf, yang memiliki arti bahwa untuk setiap nilai rataan yang tergolong dalam kelompok yang sama, perbedaan nilai rata-rata dinyatakan tidak signifikan, sementara jika berada dalam



kelompok huruf yang berbeda, sama perbedaan nilai rata-rata dinyatakan signifikan

#### **2.2.4. Radio Suara Surabaya FM**

Radio Suara Surabaya Suara Surabaya FM (SSFM) adalah sebuah stasiun radio di Kota Surabaya. Berpusat di jalan Wonokitri Besar 40, SSFM mengudara pertama kali pada tanggal 11 Juni 1983. Radio ini mengklaim sebagai radio pertama di Indonesia yang menerapkan format "radio *News*" dan informasi, dimana informasi yang sering di sampaikan antara lain kondisi lalu lintas, keamanan, dan seputar Kota Surabaya [19]. Adapun visi misi Radio Suara Surabaya FM adalah:

##### **Visi:**

SUARA SURABAYA adalah sumber pemberdayaan dan kegiatan demokratisasi masyarakat, melalui usaha kegiatan media massa yang mengikuti perkembangan teknologi komunikasi dan telekomunikasi.

##### **Misi:**

1. SUARA SURABAYA, perusahaan media massa yang dituntut berkembang dengan mengandalkan kemajuan teknologi komunikasi dan telekomunikasi.
2. SUARA SURABAYA, sentra informasi tentang Surabaya dan Jawa Timur.
3. SUARA SURABAYA menyelenggarakan berbagai kegiatan pemberdayaan proses demokratisasi masyarakat.
4. SUARA SURABAYA, sumber kehidupan dan kesejahteraan seluruh unsur karyawan yang bekerja untuk kemajuan bersama [20].

##### **2.2.4.1. Jurnalisme Warga pada Radio Suara Surabaya**

Sebagaimana peran media massa lainnya, yaitu menyajikan informasi, Radio Suara Surabaya memposisikan diri sebagai stasiun radio yang menyediakan informasi untuk warga Surabaya dan sekitarnya. Informasi yang menjadi prioritas bagi

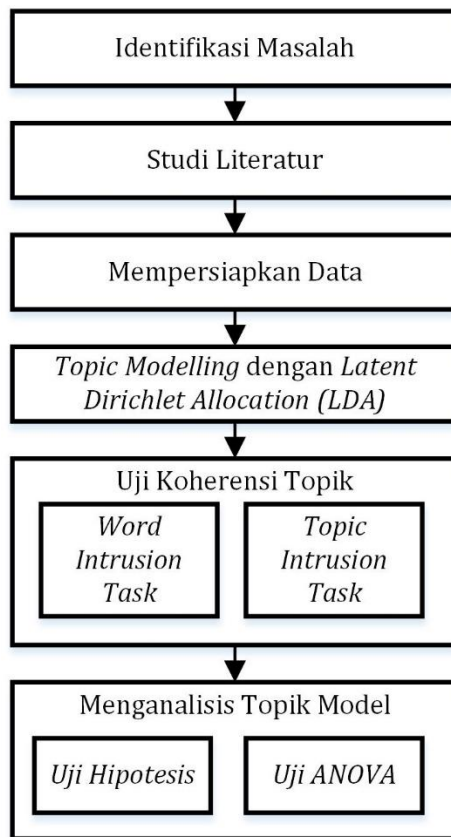
radio Suara Surabaya adalah informasi lalu lintas yang merupakan masalah besar bagi warga Surabaya. Sebagaimana media massa lainnya, Radio Suara Surabaya memiliki reporter, namun pada tahun 1994, Radio SS mengembangkan siaran interaktif berbasis *jurnalistik* masyarakat. Gagasan ini melibatkan partisipasi warga dalam melaporkan peristiwa kepada penyiar radio yang sedang bertugas [21]. Partisipasi yang melibatkan warga ini diawali berdasarkan inisiatif masyarakat sendiri, yang melaporkan kondisi lalu lintas dan cuaca yang ditanggapi positif oleh SSFM dengan menyebarkan laporan tersebut sehingga pendengar merasakan langsung manfaatnya. Menariknya SS tidak pernah mengadakan training atau menentukan format tertentu untuk pelaporan yang dilakukan masyarakat. SS memiliki penyiar radio dan *gatekeeper* yang memiliki peran untuk menyeleksi atau menyunting informasi yang disampaikan oleh masyarakat melalui media sosial [22].

#### **2.2.4.2. Media Sosial Radio Suara Surabaya**

Akun *Twitter* @e100ss dan akun *Facebook* E100 adalah akun media sosial resmi Radio Suara Surabaya sebagai salah satu saluran media komunikasi yang mendukung siaran Suara Surabaya FM (SSFM). Akun @e100ss dikembangkan Tim *On Air* Suara Surabaya pada September 2012. Nama @e100ss dipilih dengan tujuan agar mudah diingat, e menandakan *electronic*, 100 adalah Frekuensi Suara Surabaya FM dan ss merupakan singkatan dari Suara Surabaya. Sesuai visi SSFM yg mengikuti perkembangan teknologi, Suara Surabaya FM memanfaatkan media sosial @e100ss dan E100 sebagai media komunikasi resmi dengan pendengar untuk berbagi berita secara *real-time* [3].

### BAB III METODOLOGI PENELITIAN

Pada bab ini akan dijelaskan tentang metodologi yang akan digunakan dalam penyusunan tugas akhir. Metodologi akan digunakan sebagai panduan dalam penyusunan tugas akhir agar terarah dan sistematis. Adapun urutan dari pengerjaan tugas akhir dapat dilihat pada Diagram 3.1:



**Diagram 3.1 Metodologi Penelitian**

### 3.1. Identifikasi Masalah

Sebagai tahap pertama, tahap identifikasi masalah dilakukan dengan cara observasi pada media sosial informasi publik Radio Suara Surabaya FM yaitu akun *Facebook* E100 dan akun *Twitter* @e100ss. Hasil observasi menunjukkan bahwa permasalahan terdapat pada jumlah laporan masyarakat melalui media sosial yang banyak, topiknya sering berubah sewaktu-waktu dan beragam baik dari segi topik maupun format, sehingga penggalian informasi publik melalui media sosial di kota Surabaya dirasa belum optimal, padahal di kota Surabaya terdapat komunitas jurnalis warga yang aktif untuk bertukar informasi terkini dengan dipusatkan dan difasilitasi oleh Radio Suara Surabaya FM. Berdasarkan kondisi tersebut, dirasa perlu untuk melakukan identifikasi topik apa yang sedang menjadi pembicaraan pada media sosial informasi publik untuk mengetahui apa yang sedang terjadi di kota Surabaya.

### 3.2. Studi Literatur

Tahap studi literatur dilakukan dengan tujuan agar dapat memahami konsep, metode, dan teknologi sesuai bahasan dan permasalahan sehingga dapat memberi solusi mengenai permasalahan yang akan digunakan dalam penyusunan tugas akhir. Tahap studi literatur dilakukan dengan menggali informasi sesuai benang merah penelitian melalui literatur-literatur sebagai sumber terkait konsep-konsep atau penelitian sebelumnya yang pernah dilakukan yang terkait dengan permasalahan dalam bentuk jurnal, buku maupun referensi online. Adapun literatur utama pada pembahasan *Topic modeling* dan *LDA* mengacu pada penelitian David M. Blei, dkk dengan judul *Latent Dirichlet Allocation*, *Journal of Machine learning Research* 3 (2003) [4], dan pembahasan *Topic coherence* mengacu pada penelitian Jonathan Chang, dkk dengan judul *Reading Tea Leaves: How Humans Interpret Topic Models* yang terdapat pada jurnal *Neural Information Processing Systems* (2009) [18].

### 3.3. Mempersiapkan Data

Tahap mempersiapkan data terdiri dari sub aktifitas pengumpulan, pemahaman dan pemilihan data, tahap ini bertujuan untuk mempersiapkan dokumen yang akan dianalisis menggunakan *Topic modeling*. Adapun dokumen yang dianalisis adalah data media sosial milik Radio Suara Surabaya FM, yaitu akun *Facebook* E100 dan *Twitter* @e100ss yang dikumpulkan dengan cara *crawling*. Namun pada penelitian ini, *crawling* tidak termasuk dalam ruang lingkup penelitian, sehingga pengumpulan data yang dimaksud adalah menghimpun data dari berbagai hasil *crawling* yang sudah ada. Berikut merupakan hasil *crawling* akun *Twitter* @e100ss dan hasil *crawling* akun *Facebook* E100 yang disajikan pada Gambar 3.1 dan Gambar 3.2.

```
"783996019662913536", "Kejati Tetapkan Wisnu Wardhana, Mantan Ketua DPRD Surabaya Sebagai Tersangka (odp-pr)...  
https://t.co/lyvdrjjsai", "e100ss", "0.0", "0.0", "6-10-2016 18:43:11"  
"783995786707111938", "@HamidiRena @e100ss Anda benar, tapi itupun bunga yg ditawarkan masih rasional dan  
wajar bukan tanpa risiko", "0.0", "0.0", "6-10-2016 18:42:16"  
"783995712870621184", "RT @ainissfm: Margomulyo 2 arah padat lancar laporan pak Murdo' ke SS  
@e100ss", "FahmiWoles", "0.0", "0.0", "6-10-2016 18:41:58"  
"783995618351996928", "Macet di rel roomo Kalisari Surabaya @gresikNEWS @infoGRESIK @RadioElshinta @beritaKBR  
@e100ss @zainiku https://t.co/8Hcn3mltra", "bidikgresik", "0.0", "0.0", "6-10-2016 18:41:36"  
"783995505382600704", "@e100ss jalan ke arah pasar pucang dibongkar, menyebabkan macet https://t.co/vwZ3Ju9aHl  
", "nkarnadi", "0.0", "0.0", "6-10-2016 18:41:09"  
"783994435969228801", "RT @ainissfm: Margomulyo 2 arah padat lancar laporan pak Murdo' ke SS  
@e100ss", "wicakana dimas", "0.0", "0.0", "6-10-2016 18:36:54"  
"78399418966658048", "18.35: Adzan Isya' sudah berkumandang di Surabaya dan sekitarnya. Selamat menunaikan ibadah  
shalat Isya' bagi... https://t.co/0jplr2RRp7", "e100ss", "0.0", "0.0", "6-10-2016 18:35:55"  
"783993670156509184", "Merr II C adalah pilihan utk warga Rungkut dsk. Tp sebulan ini terus dikeluhkan di SS.  
MAKCETTTT. Apa perlu evaluasi? @e100ss @RTMCJatim", "ainissfm", "0.0", "0.0", "6-10-2016 18:33:51"  
"783992891496501248", "RT @e100ss: 18.26: Waspada! jalur-jalur ini.  
1. Kedungcawek arah Suramadu padat cenderung Macet imbas kecelakaan.  
2. Dinoyo... https://t.co.-", "Sosmed ntm", "0.0", "0.0", "6-10-2016 18:30:45"  
"783991745373573120", "18.26: Waspada! jalur-jalur ini.  
1. Kedungcawek arah Suramadu padat cenderung Macet imbas kecelakaan.  
2. Dinoyo... https://t.co/ZvIKf0Gqia", "e100ss", "0.0", "0.0", "6-10-2016 18:26:12"  
"783991661621719040", "RT @teguhaw79: @e100ss PJU mati Gelapnya desa bakalan rt 4-9 balongbendo sdh tip DPK ndak  
ada tindak lanjut, harus lapor kmm lagi?masakhrs..", "e100ss", "0.0", "0.0", "6-10-2016 18:25:52"  
"783991594886111232", "@e100ss PJU mati Gelapnya desa bakalan rt 4-9 balongbendo sdh tip DPK ndak ada tindak  
lanjut, harus lapor kmm lagi?masakhrs Bupati sidoarjo", "teguhaw79", "0.0", "0.0", "6-10-2016 18:25:36"
```

Gambar 3.1 Hasil *Crawling* Akun *Twitter* @e100ss

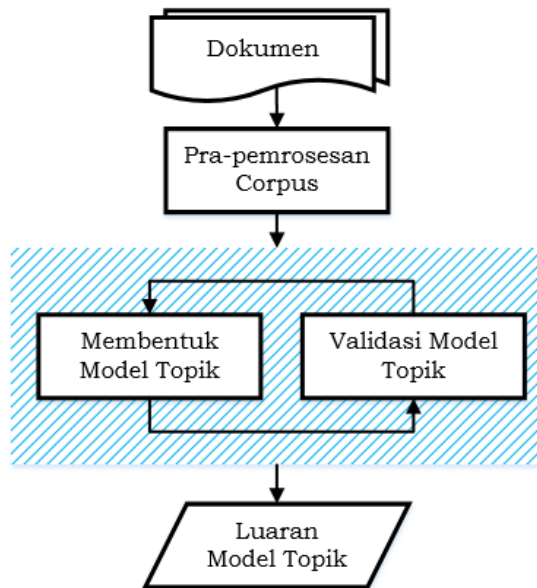
```
"227268729878_10153884281089879","20.50 : Hindari masuk JL Kalidami! JL Kalidami - Unair Kampus B ada bazar. Lalu
lintas MACET Karena jalur yang dari arah Karang Menjangan digunakan jadi 2 lajur. (odp-rt)","", "1-8-2015 20:51:23"
"227268729878_10153884250709879","Masih Ada Pilkada Parlon Tunjpal, Tahapan Pilkada Lanjut Terus. (odp-rt)
http://m.suarasurabaya.net/politik/detail.php?id=k878thocermm5g8a7r6esbdr72015156515","", "1-8-2015 20:34:02"
"227268729878_10153884236539879","#SSinfo : Unjuk rasa ribuan karyawan Migas Blok Cepu Bojonegoro ricuh, Sabtu
(18/8/2015). Nana reporter Radio Suara Bojonegoro Indah melaporkan ribuan massa merusak kantor dan pos security.
Empat Mobil digulung dan satu mobil dibakar. Unjuk rasa menuntut dibukanya kembali lima pintu utama masuk dan
keluar karyawan. Saat ini, hanya satu pintu yang digunakan. Jadi karyawan harus berdesak-desakan untuk
keluar,masuk dan istirahat. Tidak adanya jawaban dari pihak perusahaan memicu timbulnya keributan. Imbas kejadian
ini, kegiatan perusahaan diliburkan sampai batas waktu yang tidak bisa ditentukan. Foto : Dokumentasi Radio
Suara Bojonegoro Indah. (odp-rt)","E100 added 2 new photos.", "1-8-2015 20:22:57"
"227268729878_1015388479199879","19.53 : Info awal : Kecelakaan di Jalang FO Peterongan Jombang antara mobil
dengan sepeda motor. Lalu lintas MACET. Belum ada polisi dilokasi. (odp-rt)","", "1-8-2015 19:54:26"
"227268729878_10153884123594879","19.38 : Jalur luar kota MACET :
1.Balong Bendo - Mojokerto;
2.Peterongan Jombang - Surabaya. (odp-rt)","", "1-8-2015 19:40:14"
"227268729878_10153884087459879","19.15 : 3 Jalur MACET malam ini:
1.Jembatan Karangpilang Baru dan lama - Sepanjang;
2.Kletek - Krian. Ada pick up mogok didepan SFBU Kletek;
3.Mojoaung. (odp-rt)","", "1-8-2015 19:15:41"
"227268729878_10153884045209879","19.53 : Info awal : Kebakaran lahan kosong di depan Lenmarc. Api membesar. PKK
sudah menuju lokasi. Foto : Emig Mia Via @100. (odp-rt)","", "1-8-2015 19:54:35"
"227268729878_10153884047289879","18.45 : Adzan Isya telah berkumandang untuk wilayah Surabaya dan sekitarnya.
Selamat menunaikan ibadah sholat Isya, Kawan. (odp-rt)","", "1-8-2015 18:46:52"
```

**Gambar 3.2 Hasil Crawling Akun Facebook E100**

Berdasarkan data hasil *crawling* tersebut, sebagai contoh hasil *crawling* pada *Twitter*, data tersebut mengandung atribut "id", "message", "account", "latitude", "longitude", dan "created\_time", namun tidak seluruh atribut data digunakan untuk *Topic modeling*, sehingga diperlukan tahap pemilihan data. Adapun pemilihan data yang dimaksud adalah pemilihan kolom apa saja yang selanjutnya akan dianalisis. Dalam penelitian ini, kolom yang digunakan terbatas pada isi pesan (message) baik pada *Twitter* maupun *Facebook*, yang mana setiap baris data dapat diartikan sebagai satu dokumen. Apabila diperhatikan lebih lanjut, data yang tersimpan dalam isi pesan (message) perlu dilakukan pembersihan data terlebih dahulu, seperti penghapusan link, *username* maupun data lain yang dianggap dapat mengganggu proses-proses berikutnya.

### 3.4. *Topic modeling dengan Latent Dirichlet Allocation*

Tahap *Topic modeling* dengan *Latent Dirichlet Allocation (LDA)* terhadap dokumen yang berasal dari media sosial *Facebook* dan *Twitter* resmi Radio Suara Surabaya FM terdiri dari beberapa tahap, yang digambarkan melalui Diagram 3.2



**Diagram 3.2 Tahap *Topic modeling* dengan *Latent Dirichlet Allocation (LDA)***

### **3.4.1. Pra-pemrosesan *Corpus***

Dalam melakukan *Topic modeling* dengan *LDA*, diperlukan langkah-langkah untuk mempersiapkan data sehingga dapat diolah pada tahap berikutnya, tahap ini disebut tahap pra-pemrosesan *corpus*. Adapun sub-aktivitas dari tahap pra-pemrosesan *corpus* digambarkan dengan Diagram 3.3

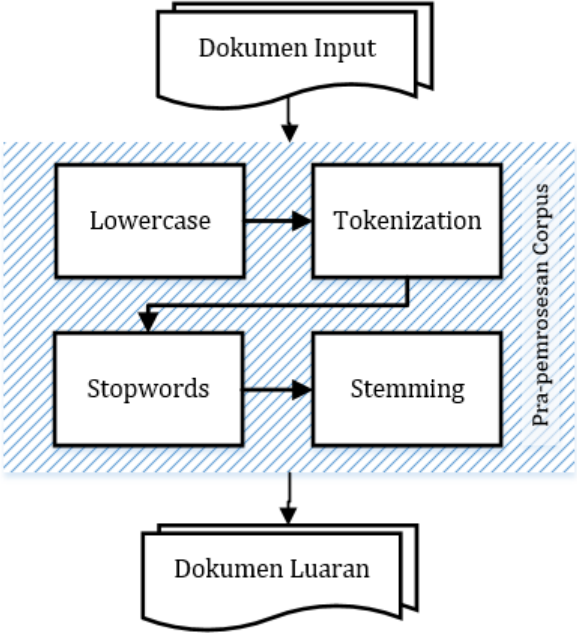


Diagram 3.3 Sub-aktivitas dari tahap pra-pemrosesan *corpus*

Data teks perlu dibentuk menjadi menjadi *lowercase* dengan tujuan agar kata yang sama namun berbeda secara penulisan huruf kapital dan tidak, tidak dianggap kata yang berbeda. Contoh data sebelum dan sesudah melalui tahap *lowercase* ditampilkan pada Tabel 3.1

Tabel 3.1 Contoh Data Sebelum dan Sesudah Melalui Tahap *Lowercase*

Sebelum <i>Lowercase</i>	Setelah <i>Lowercase</i>
Perampasan            Aset	perampasan            aset
Century di    Hongkong	century di    hongkong
Butuh            Proses	butuh            proses
Perampasan aset Bank	perampasan aset bank
Century    (kini    Bank	century    (kini    bank
Mutiara) di    Hongkong	mutiara) di    hongkong



membutuhkan proses meskipun fatwa dari Mahkamah Agung telah dijadikan dasar Kejaksaan Agung Hingga saat ini Kejaksaan Agung masih menunggu fatwa khusus yang akan diterbitkan MA	membutuhkan proses meskipun fatwa dari mahkamah agung telah dijadikan dasar kejaksaan agung hingga saat ini kejaksaan agung masih menunggu fatwa khusus yang akan diterbitkan ma
---	---

*Tokenization* adalah aktivitas atau proses memisahkan deretan kata di dalam kalimat atau paragraf menjadi potongan kata tunggal atau *termmed word*. Proses *tokenization* bertujuan untuk mempersiapkan dokumen untuk proses berikutnya, yaitu proses *stopwords* dan *Stemming* dapat dilakukan. Contoh data sebelum dan sesudah melalui tahap *Tokenization* ditampilkan pada Tabel 3.2

**Tabel 3.2 Contoh Data Sebelum dan Sesudah Melalui Tahap Tokenization**

<b>Sebelum Tokenization</b>	<b>Setelah Tokenization</b>
perampasan aset century di hongkong butuh proses perampasan aset bank century (kini bank mutiara) di hongkong membutuhkan proses meskipun fatwa dari mahkamah agung telah dijadikan dasar kejaksaan agung hingga saat ini kejaksaan agung masih menunggu fatwa khusus yang akan diterbitkan ma	['perampasan', 'aset', 'century', 'di', 'hongkong', 'butuh', 'proses', 'perampasan', 'aset', 'bank', 'century', 'kini', 'bank', 'mutiara', 'di', 'hongkong', 'membutuhkan', 'proses', 'meskipun', 'fatwa', 'dari', 'mahkamah', 'agung', 'telah', 'dijadikan', 'dasar', 'kejaksaan', 'agung', 'hingga', 'saat', 'ini',

	'kejaksaan', 'agung', 'masih', 'menunggu', 'fatwa', 'khusus', 'yang', 'akan', 'diterbitkan', 'ma']
--	--

*Stopwords* merupakan kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh kelas kata termasuk *stopwords* ditunjukkan dengan Tabel 3.3

**Tabel 3.3 Contoh Kelas Kata**

<b>Kelas Kata</b>	<b>Contoh</b>
Kata depan	di, pada, dari, ke, kepada, akan, oleh, daripada, hingga, sampai, dll
Kata Sambung	dan, dengan, serta atau, sedangkan, sebab, jika, bila, sebagai, sehingga, sesudah, dll
Kata Ganti	saya, aku, ku, kami, kita, kamu, engkau, anda, kalian, ia, dia, beliau, mereka, dll

Daftar *stopwords* yang digunakan adalah *stopwords* Bahasa Indonesia yang disusun berdasarkan penelitian Fadillah Z Tala [23]. Beberapa contoh daftar *stopwords* yang tersimpan dalam daftar yang dimaksud tercantum pada Tabel 3.4

**Tabel 3.4 Contoh daftar *stopwords* berdasarkan penelitian Fadillah Z Tala [23]**

<b>Daftar <i>Stopwords</i></b>	
'ada',	'agak',
'bahwa',	'cuma',
'dapat',	'entah',
'ialah',	'ini',
'lalu',	'mana',
'pasti',	'para',
'tiap',	'yakni',
'amatlah',	'bagaimana',
'cukup',	'demikian',
'guna',	'hal',
'jika',	'juga',
'kiranya',	'oleh',
'memang',	'namun',
'saat',	'sangat',
'tanpa',	'yaitu'

Menghilangkan *stopwords* merupakan tahap yang penting, mengingat tingginya frekuensi kemunculan *stopwords* dalam dokumen, yang berujung pada tingginya probabilitas kata-kata *stopwords* dalam topik, sehingga topik tidak dapat diinterpretasi dengan baik. Contoh data sebelum dan sesudah melalui tahap penghapusan *Stopwords* ditampilkan pada Tabel 3.5

**Tabel 3.5 Contoh Data Sebelum dan Sesudah Melalui Tahap Penghapusan *Stopwords***

<b>Sebelum Penghapusan <i>Stopwords</i></b>	<b>Setelah Penghapusan <i>Stopwords</i></b>
'perampasan', 'aset',	['perampasan',
'century', 'di',	'aset', 'century', ,
'hongkong', 'butuh',	'hongkong', 'butuh',
'proses',	'proses',
'perampasan', 'aset',	'perampasan', 'aset',
'bank', 'century',	'bank', 'century',
'kini', 'bank',	'bank', 'mutiara',
'mutiara', 'di',	'hongkong',
'hongkong',	'membutuhkan',
'membutuhkan',	'proses', 'fatwa',

'proses', 'meskipun',	'mahkamah', 'agung',
'fatwa', 'dari',	'dijadikan', 'dasar',
'mahkamah', 'agung',	'kejaksaan', 'agung',
'telah', 'dijadikan',	'hingga',
'dasar', 'kejaksaan',	'kejaksaan', 'agung',
'agung', 'hingga',	'menunggu', 'fatwa',
'saat', 'ini',	'khusus',
'kejaksaan', 'agung',	'diterbitkan', 'ma']
'masih', 'menunggu',	
'fatwa', 'khusus',	
'yang', 'akan',	
'diterbitkan', 'ma'	

*Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut. *Stemming* bekerja dengan menghilangkan semua imbuhan (*affixes*), baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan kombinasi dari awalan dan akhiran (*confixes*) pada kata turunan. Data teks perlu dibentuk menjadi kata dasarnya dengan tujuan agar tidak terdapat kata yang sama namun berbeda karena adanya imbuhan (*affixes*). Adapun proses *Stemming* dilakukan dengan menggunakan *library* Sastrawi, yaitu *library Stemmer* bahasa indonesia dengan lisensi MIT yang memanfaatkan kamus kata dasar dari Kateglo sebagai acuan. Contoh data sebelum dan sesudah melalui tahap *Stemming* ditampilkan pada Tabel 3.6

Tabel 3.6 Contoh Data Sebelum dan Sesudah Melalui Tahap *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
'perampasan', 'aset',	['ampas', 'aset',
'century',	'century',
'hongkong', 'butuh',	'hongkong', 'butuh',
'proses',	'proses', 'ampas',
'perampasan', 'aset',	'aset', 'bank',
'bank', 'century',	'century', 'bank',
'bank', 'mutiara',	'mutiara',
'hongkong',	'hongkong', 'butuh',
'membutuhkan',	'proses', 'fatwa',
'proses', 'fatwa',	'mahkamah', 'agung',

'mahkamah',	'agung',	'jadi',	'dasar',
'dijadikan',	'dasar',	'jaksa',	'agung',
'kejaksaan',	'agung',	'jaksa',	'agung',
'hingga',	'kejaksaan',	'tunggu',	'fatwa',
'agung',	'menunggu',	'khusus',	'terbit',
'fatwa',	'khusus',	'ma']	
'diterbitkan',	'ma'		

### 3.4.2. Membentuk Model Topik

Tahap membentuk topik model bertujuan untuk menghasilkan model topik yang paling tepat untuk dokumen. Model topik dikatakan tepat apabila mampu menghasilkan luaran yang baik pada tahap validasi model topik. Untuk menghasilkan model topik yang tepat, hal yang dilakukan adalah dengan melakukan eksperimen pada nilai *input parameter*. Adapun parameter yang dimaksud adalah *number of topics* dan *words in topic*. Parameter *number of topics* menentukan jumlah topik dalam satu dokumen, sementara parameter *number of words in topic* menunjukkan jumlah kata penyusun topik. Berikut merupakan contoh luaran model topik yang dilakukan dengan eksperimen *input parameter number of topics* sejumlah 30 dan *words in topic* sejumlah 6 yang ditunjukkan pada Tabel 3.7

**Tabel 3.7 Contoh Luaran Model Topik Hasil Eksperimen**

Topik	Probabilitas*Kata
0	0.020*indonesia + 0.017*harga + 0.007*rumah + 0.006*bahan + 0.006*singapura + 0.006*minyak
1	0.009*ipad + 0.008*orang + 0.007*jual + 0.006*beli + 0.006*minyak + 0.006*harga
2	0.023*rumah + 0.009*harga + 0.009*indonesia + 0.008*jual + 0.008*unit + 0.008*milik
3	0.033*film + 0.022*balap + 0.012*dunia + 0.011*seri + 0.011*juara + 0.008*musim
4	0.033*didik + 0.028*sekolah + 0.027*siswa + 0.024*guru + 0.020*ajar + 0.018*anak
5	0.017*main + 0.015*unggul + 0.014*tenis + 0.012*juara + 0.011*raih + 0.009*babak
6	0.009*israel + 0.009*dapur + 0.006*negara + 0.005*perintah + 0.005*keramik + 0.005*indonesia

7	0.031*lagu + 0.016*musik + 0.012*band + 0.012*album + 0.009*nyanyi + 0.009*aku
8	0.009*surya + 0.009*temu + 0.009*beasiswa + 0.008*favorit + 0.007*layan + 0.006*film
9	0.009*ruang + 0.008*kpk + 0.006*bangun + 0.006*tahan + 0.006*menteri + 0.005*rumah
10	0.014*pariwisata + 0.008*indonesia + 0.006*wisata + 0.006*tingkat + 0.005*wisatawan + 0.005*promosi
11	0.014*indonesia + 0.008*persen + 0.008*rp + 0.007*mahasiswa + 0.006*orang + 0.005*juta
12	0.012*risiko + 0.008*kanker + 0.008*sel + 0.008*stroke + 0.007*teliti + 0.006*kulit
13	0.010*indonesia + 0.010*negara + 0.006*hasil + 0.005*milu + 0.005*kpk + 0.005*komisi
14	0.008*rp + 0.007*as + 0.007*suriah + 0.006*indonesia + 0.006*laba + 0.005*pohon
15	0.020*mobil + 0.010*motor + 0.009*kendara + 0.008*mesin + 0.006*cepat + 0.005*honda
16	0.010*temu + 0.009*nama + 0.008*laba + 0.007*gajah + 0.007*makan + 0.006*ilmuwan
17	0.009*mars + 0.008*robot + 0.006*saham + 0.006*spirit + 0.005*ubah + 0.005*tipe
18	0.027*sakit + 0.019*obat + 0.015*jantung + 0.012*sebab + 0.010*serang + 0.009*dokter
19	0.008*game + 0.007*galaksi + 0.007*kalah + 0.007*alpha + 0.007*kamera + 0.006*fitur
20	0.015*makan + 0.011*sakit + 0.010*diabetes + 0.009*darah + 0.009*sehat + 0.009*anak
21	0.013*ponsel + 0.011*android + 0.011*guna + 0.008*aplikasi + 0.007*indonesia + 0.007*google
22	0.009*presiden + 0.008*perintah + 0.007*indonesia + 0.006*temu + 0.005*jalan + 0.005*vila
23	0.009*indonesia + 0.008*negara + 0.007*budaya + 0.006*bukti + 0.005*batak + 0.005*pasang
24	0.015*bangun + 0.009*kembang + 0.009*wisata + 0.008*persen + 0.007*kota + 0.007*perintah
25	0.009*as + 0.009*pasu + 0.009*afghanistan + 0.008*sehat + 0.008*serang + 0.006*hasil
26	0.009*film + 0.008*kaki + 0.008*microsoft + 0.007*juang + 0.007*partai + 0.007*kader
27	0.010*indonesia + 0.008*wisata + 0.007*kawasan + 0.006*jakarta + 0.006*perintah + 0.005*nasional
28	0.017*batu + 0.012*china + 0.008*ekspor + 0.008*jenis + 0.007*film + 0.007*telepon
29	0.015*persen + 0.015*turun + 0.013*saham + 0.011*rp + 0.008*indonesia + 0.008*as

### 3.4.3. Validasi Model Topik

Tahap validasi topik bertujuan untuk memastikan model topik yang dihasilkan dari hasil *Topic modeling* yang dilakukan pada dokumen adalah benar, baik luaran berupa topik maupun kata-kata dalam topik. Dalam hal ini, tingkat kebenaran topik disesuaikan dengan dua metode, secara otomatis dengan *Perplexity* dan berdasarkan tingkat koherensi atau tingkat kemudahannya dalam diinterpretasi manusia.

Ide dasar *Perplexity* adalah membagi secara acak dokumen menjadi data *training* dan data uji, kemudian menghitung rata-rata probabilitas *log* setiap kata dari data uji pada model yang dihasilkan dari data *training* [24], atau secara sederhana, *perplexity* dilakukan dengan mengambil  $n$  sampel dari  $N$  populasi data untuk diuji, apakah  $n$  sampel tersebut memiliki kesesuaian topik dengan kelompok topik dalam  $N$  populasi. Secara otomatis, penghitungan *Perplexity* sudah termasuk dalam *package gensim* untuk bahasa *python*, cara kerjanya adalah dengan menghitung rata-rata jarak geometris dari matriks data yang mewakili setiap kata menggunakan potongan dokumen evaluasi *corpus*. Apabila nilai *Perplexity* dirasa sudah mencapai titik optimal, maka akan dilakukan validasi model topik berdasarkan tingkat koherensinya sementara apabila model topik belum mencapai nilai *perplexity* yang optimal, maka akan kembali ke tahap membentuk topik model untuk kembali dilakukan eksperimen pada *input parameter*.

### 3.5. Uji Koherensi Topik

Untuk validasi topik berdasarkan tingkat koherensi atau tingkat kemudahannya dalam diinterpretasi manusia, validasi topik dilakukan dengan metode *Topic Intrusion task* dan validasi kata-kata dalam topik dilakukan dengan *Word Intrusion Task*. *Word Intrusion task* dilakukan dengan menyisipkan sebuah kata dengan probabilitas rendah dalam kata-kata hasil proses *Topic modeling* pada suatu topik tertentu, kemudian meminta beberapa responden untuk menebak manakah kata yang bukan hasil proses *Topic modeling*. *Topic*

*Intrusion task* dilakukan dengan menyajikan cuplikan dokumen yang disertai beberapa pilihan topik kepada beberapa responden, yang mana salah satu dari beberapa topik tersebut merupakan topik yang memiliki probabilitas rendah dalam dokumen.

Mengingat tahap validasi topik ini bertujuan untuk memastikan luaran dari hasil *Topic modeling* yang dilakukan pada dokumen memiliki koherensi yang baik, maka apabila topik belum memiliki tingkat koherensi yang optimal, maka akan kembali ke tahap membentuk topik model untuk kembali dilakukan eksperimen pada *input parameter*.

### **3.6. Menganalisis Topik Model**

Tahap menganalisis topik model bertujuan untuk mendapatkan kesimpulan terkait hasil uji koherensi topik secara lebih spesifik. Menganalisis topik model dilakukan dengan melakukan uji hipotesis dan uji ANOVA. Uji hipotesis merupakan pengujian yang dilakukan pada tingkat metode dan pengaruh perlakuan. Uji hipotesis dilakukan dengan melakukan pembuktian terhadap hipotesis yang telah ditentukan sebelumnya dengan metode statistika sehingga didapatkan kesimpulan terkait hasil perbandingan antara word intrusion task dengan topic intrusion task dan perbandingan hasil pengaruh perlakuan stem. Uji ANOVA bertujuan untuk mengetahui derajat kemudahan interpretasi atau pemahaman responden pada tingkat topik yang dihasilkan. Uji ANOVA dilakukan dengan melakukan perbandingan secara berpasangan sehingga didapatkan kelompok topik sesuai kemudahannya untuk diinterpretasi oleh manusia untuk setiap task pada uji koherensi topik



## BAB IV PERANCANGAN

Pada bab ini akan dijelaskan bagaimana rancangan dari penelitian tugas akhir yang meliputi subyek dan obyek dari penelitian, pemilihan subyek dan obyek penelitian dan bagaimana penelitian akan dilakukan.

### 4.1. Pengambilan Data

Dalam pelaksanaan analisis informasi topik informasi publik media sosial di Surabaya, data merupakan objek utama analisis. Data yang dibutuhkan merupakan data tipe teks hasil *crawling* dari akun media sosial resmi Radio Suara Surabaya. Adapun data yang akan diolah adalah *post* FB akun E100 termasuk komentar-komentarnya, dan *tweet* akun @e100ss dan *tweet* yang *me-mention* akun @e100ss.

Pengambilan data dilakukan dengan melakukan *crawling* data pada *post* FB akun E100 termasuk komentar-komentarnya, dan *tweet* akun @e100ss dan *tweet* yang *me-mention* akun @e100ss. Proses *crawling* menghasilkan data pesan informasi publik yang dipublikasi via media sosial dari bulan September 2016 hingga bulan Januari 2017. Adapun data pesan informasi publik tersebut disimpan dalam tiga tabel dalam format *.sql* dengan *database management syStem (DBMS) MySQL* yang diuraikan pada Tabel 4.1:

**Tabel 4.1 Tabel Sumber Data Masukan**

Nama Tabel	Keterangan
<i>fb_test5</i>	merupakan tabel yang menyimpan data pesan informasi publik yang di- <i>post</i> melalui akun <i>Facebook</i> resmi Radio Suara Surabaya
<i>fb_comments</i>	merupakan tabel yang menyimpan data komentar pada pesan informasi publik yang

	di-post melalui akun <i>Facebook</i> resmi Radio Suara Surabaya
<i>tw_test2</i>	merupakan tabel yang menyimpan data pesan informasi publik yang di-tweet melalui akun <i>Twitter</i> resmi Radio Suara Surabaya dan tweet dari masyarakat yang me-mention akun e100ss

Atribut pada tabel fb\_test5 dijelaskan pada Tabel 4.2

**Tabel 4.2 Tabel Atribut pada Tabel fb\_test5**

Nama Atribut	Tipe Data	Keterangan
<i>fb_id</i>	<i>text</i>	merupakan kode unik untuk mengidentifikasi akun pengguna <i>Facebook</i> yang mengirim pesan
<i>message</i>	<i>text</i>	merupakan pesan yang dikirim
<i>created_time</i>	<i>datetime</i>	merupakan catatan waktu kapan pesan dikirim
<i>user</i>	<i>text</i>	merupakan akun yang mengirim pesan terkait

Contoh data pada tabel fb\_test5 dapat dilihat pada Gambar 4.1

fb_id	message	story	created_time	user
227268729878_10153884619194879	Muktamar NU rumuskan konsep Islam Nusantara. Islam.	E100 with Naira Purnomo	2015-08-01 23:41:38	e100ss
227268729878_10153884589739879	Guyonan Gus Ipul, Pakde Kanro dan Jokowi di Muktam.	E100 with Naira Purnomo	2015-08-01 23:27:21	e100ss
227268729878_10153884372594879	Sebanyak 94 TKI ilegal Dideportasi Malaysia (odp-		NULL	2015-08-01 22:10:00
227268729878_10153884372739879	Datang ke Muktamar, Jokowi Bagikan Kaos dan Kartu		NULL	2015-08-01 21:55:31
227268729878_10153884360259879	21.45 - Hindari masuk Jombang Kotal Lalu lintas MA.		NULL	2015-08-01 21:46:27
227268729878_10153884312839879	Foto almarhum KH Abdurrahman Wahid alis Gus Dur se.	E100 with Imam De Ives and Naira Purnomo	2015-08-01 21:16:06	e100ss
227268729878_10153884298174879	21.00 : 4 Jalur MACET : 1 Simpang 3 Laksantri.		NULL	2015-08-01 21:04:19
227268729878_10153884281089879	20.50 - Hindari masuk Jl. Kalidami/ Jl. Kalidami - U.		NULL	2015-08-01 20:51:23
227268729878_10153884250709879	Masih Ada Pilkada Paslon Tunggal. Tahapan Pilkada ...		NULL	2015-08-01 20:34:02
227268729878_10153884236539879	#SSinfo : Unjuk rasa ribuan karyawan migas Blok Co.	E100 added 2 new photos.	2015-08-01 20:22:57	e100ss

**Gambar 4.1 Contoh data pada tabel fb\_test5**

Sementara atribut pada data komentar pada tabel fb\_comments dijelaskan pada Tabel 4.3

**Tabel 4.3 Tabel Atribut pada Tabel fb\_comments**

Nama Atribut	Tipe Data	Keterangan
--------------	-----------	------------

<i>id_comment</i>	<i>text</i>	merupakan kode unik untuk megidentifikasi komentar yang dikirim
<i>id_post</i>	<i>text</i>	merupakan kode unik untuk megidentifikasi pesan ( <i>posting</i> ) yang dikirim
<i>message</i>	<i>text</i>	merupakan komentar yang dikirim pada pesan terkait
<i>account</i>	<i>text</i>	merupakan akun yang mengirim komentar
<i>account_id</i>	<i>text</i>	merupakan kode unik untuk megidentifikasi akun yang mengirim komentar
<i>created_time</i>	<i>text</i>	merupakan catatan waktu komentar dikirim

Contoh data pada tabel fb\_comments dapat dilihat pada Gambar 4.2

m	id_comment	id_post	message	account	account_id	created_time
NULL	10153884619194879_10153885592159879	227268729878_10153884619194879	islam yang mau kongsi dg partai sekuler dan anti r	Aryuni Mas'udah	10207712766760292	2015-08-02 06:55:23
NULL	10153884619194879_1015388559684879	227268729878_10153884619194879	Beda islam nusantara dg islam yg sbenera nabi muham	Choirul Wachid Syamsudin	1316568895033631	2015-08-02 06:37:20
NULL	10153884619194879_10153885605864879	227268729878_10153884619194879	itu konsep yg bagus... ISLAM NUSANTARA. bnp pent...	Hady Sudono	1433513476677736	2015-08-02 07:02:40
NULL	10153884619194879_10153885156314879	227268729878_10153884619194879	Miris sekali Baca komen2x sukses betul para pem	Rudi Suaminya Farida	1378714102146405	2015-08-02 03:07:51
NULL	10153884619194879_10153885225279879	227268729878_10153884619194879	Opo maneh islam nusantara... adanya hanya ISLAM.	Erlanmilla Puspitasari	10207103612895392	2015-08-02 04:01:30
NULL	10153884619194879_10153885507539879	227268729878_10153884619194879	islam dangdutan. Wkwkkk	Yitno Wijaya	10207539940719118	2015-08-02 06:00:26
NULL	10153884619194879_10153885573899879	227268729878_10153884619194879	Berarti dalam pemahaman NU orang yang islam saja	Indro Gunawan	10207262676887140	2015-08-02 06:47:34
NULL	10153884619194879_10153885497764879	227268729878_10153884619194879	Wkwkkk... islam nusantara itu gak ada. itu hanya ak...	Hidayatun Rohman	1322753934484107	2015-08-02 05:52:17
NULL	10153884619194879_101538856180054879	227268729878_10153884619194879	Pernah waktu ngaji dulu diucapkan... bla sesora...	Muhammad Izzul	853171324819275	2015-08-02 07:10:18
NULL	10153884619194879_10153885471614879	227268729878_10153884619194879	islam cuma 1... iki opo maneh islam nusantara? Inj...	Milu Rakhmanu	1916134601947458	2015-08-02 05:46:17

**Gambar 4.2** Contoh data pada tabel fb\_comments

Sedangkan atribut pada data tabel tw\_test2 dijelaskan pada Tabel 4.4

**Tabel 4.4** Tabel Atribut pada Tabel tw\_test2

Nama Atribut	Tipe Data	Keterangan
<i>id</i>	<i>text</i>	merupakan kode unik untuk megidentifikasi <i>tweet</i> yang dikirim

<i>message</i>	<i>text</i>	merupakan <i>tweet</i> yang dikirim
<i>account</i>	<i>text</i>	merupakan akun yang mengirim <i>tweet</i> terkait
<i>latitude</i>	<i>text</i>	merupakan lokasi dimana akun mengirim <i>tweet</i> terkait
<i>longitude</i>	<i>text</i>	
<i>created_time</i>	<i>datetime</i>	merupakan catatan waktu <i>tweet</i> dikirim

Contoh data pada tabel *tw\_test2* dapat dilihat pada Gambar 4.3

id	message	account	latitude	longitude	created_time
99929831344582656	@masinusina saya maju pilgub DKI melalui jalur "in...	basuki_btp	0.0	0.0	2011-08-07 02:48:23
99626186384224257	Sudah pasti! RT @ulanadzani: Dukung pa @basuki_btp...	basuki_btp	0.0	0.0	2011-08-06 06:41:48
99625836428267520	Terima kasih RT @ZeffryAlamsyah: Mudah2an pak Ahok...	basuki_btp	0.0	0.0	2011-08-06 06:40:25
99624089026707457	Laporan kegiatan reses ke Bangka bgn 1. http://bit...	basuki_btp	0.0	0.0	2011-08-06 06:33:28
99512903534784512	@IndahS99 Terimakasih atas masukannya. Setidaknya ...	basuki_btp	0.0	0.0	2011-08-05 23:11:39
99511454633439232	Semua solusi utk masalah DKI sdh ada. Yg dibutuhka...	basuki_btp	0.0	0.0	2011-08-05 23:05:54
99510465989836800	Modal ratusan milyar ngga punya. Hanya BTP & hati ...	basuki_btp	0.0	0.0	2011-08-05 23:01:58
99509268797063168	Semua solusi utk masalah DKI sdh ada. Selama ini n...	basuki_btp	0.0	0.0	2011-08-05 22:57:13
99504618605785088	@ARIEFSYAIFUL Saya setuju. Terima kasih atas masuk...	basuki_btp	0.0	0.0	2011-08-05 22:38:44
99497933988311040	@ARIEFSYAIFUL Hrsnya demikian, termasuk di kement...	basuki_btp	0.0	0.0	2011-08-05 22:12:10

**Gambar 4.3** Contoh data pada tabel *tw\_test2*

## 4.2. Observasi Data Mentah

Observasi data mentah merupakan tahapan yang dilakukan untuk memahami data yang digunakan sebagai objek analisis. Berdasarkan pengamatan yang dilakukan, data pesan media sosial memiliki sifat yang sangat beragam dari kandungan informasinya, sehingga perlu dilakukan justifikasi manakah data yang bersifat informatif dan manakah data yang non-informatif atau *casual communication*. Adapun hasil observasi data dan justifikasi yang dilakukan tercantum pada Tabel 4.5

**Tabel 4.5** Justifikasi data masukan

Jenis Data	Sifat Data	Justifikasi	Keterangan
<i>Post</i> berita dari akun <i>Facebook</i>	informatif	merupakan pesan terkait berita-berita yang sedang	Digunakan untuk tahap selanjutnya. Pesan berita

		terjadi di masyarakat	melalui akun <i>Facebook</i> dipilih karena mampu mengakomodasi lebih banyak karakter sehingga berita yang dikirim lebih lengkap
komentar dari pesan yang dikirim akun <i>Facebook</i> E100	non-informatif	merupakan komentar yang disampaikan masyarakat pada pesan yang dikirim akun <i>Facebook</i> E100	Tidak digunakan karena tidak memiliki dasaran yang jelas
<i>Tweet</i> berita dari akun e100ss	informatif	merupakan <i>tweet</i> terkait berita-berita yang sedang terjadi di masyarakat. Pesan berita ini juga disampaikan pada akun <i>Facebook</i> E100	tidak digunakan karena redundan dengan pesan berita yang dikirim melalui akun <i>Facebook</i> .
<i>Tweet</i> masyarakat yang <i>mention</i> e100ss	campuran antara data informatif dan non-informatif	merupakan pesan dari masyarakat yang diperuntukkan	yang digunakan adalah pesan dari masyarakat yang di- <i>tweet</i> kembali oleh

		kepada Radio Suara Surabaya. Sangat beragam dari tingkat informasi yang disampaikan sehingga perlu disaring agar didapatkan pesan dengan tingkat informasi yang tinggi.	akun e100ss, karena merupakan pesan yang dengan pertimbangan tertentu dianggap informatif oleh <i>gatekeeper</i> e100ss sehingga dirasa perlu untuk <i>directtweet</i>
--	--	---	--

Berdasarkan justifikasi tersebut, data yang akan digunakan pada penelitian ini adalah pesan yang dinilai memiliki kandungan informasi yang baik, yaitu pesan berita yang dikirim melalui akun *Facebook* E100 dan *tweet* dari masyarakat yang di-*tweet* kembali oleh akun e100ss.

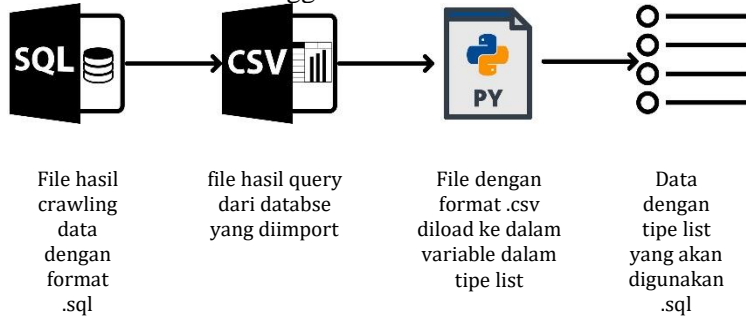
### 4.3. Metodologi Implementasi Penelitian

Metodologi implementasi penelitian merupakan tahapan yang dilakukan untuk mencapai tujuan penelitian yang disesuaikan dengan komputasi secara otomatis. Ada 5 tahapan utama dalam melakukan implementasi penelitian yaitu mempersiapkan data, pra-proses data, *Topic modeling* dengan *LDA*, uji koherensi topik, dan menganalisis topik model

#### 4.3.1. Mempersiapkan Data

Tahap mempersiapkan data merupakan tahap untuk mengolah data sedemikian rupa agar siap sebelum melakukan analisis pada penelitian. Data mentah dalam format .sql yang diolah dalam database management sy~~St~~em (DBMS) MySQL diubah menjadi format .csv agar lebih mudah untuk diolah. Data

yang dimuat adalah data keseluruhan suara pelanggan dari bulan Oktober 2015 hingga bulan Januari 2017.

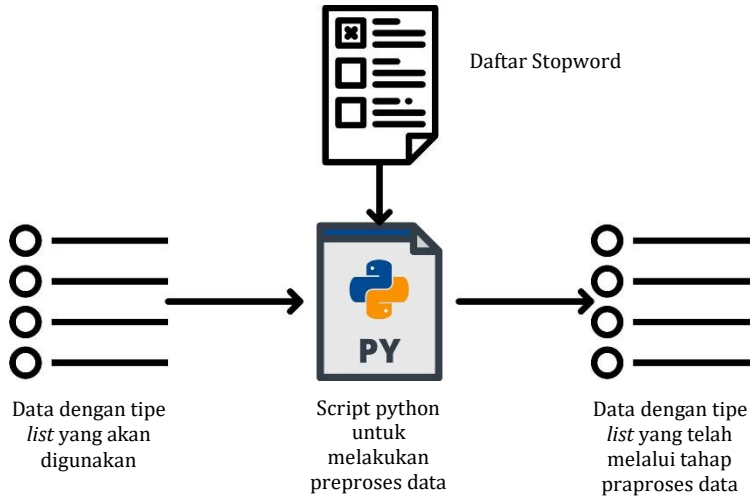


**Diagram 4.1 Alur mempersiapkan data**

#### 4.3.2. Pra-Proses Data

Pra-proses data mencakup beberapa langkah utama pengerjaan yakni membersihkan data dan *stopword* removal. Pembersihan data dilakukan untuk menghilangkan penulisan huruf besar menjadi huruf kecil, menghilangkan adanya karakter yang tidak diperlukan, menghapus kata dengan jumlah huruf kurang dari 2 dan menghapus token yang hanya terdiri dari angka.

Pendefinisian *stopword* mengacu pada susunan *stopword* yang telah disusun oleh penelitian Fadillah Z Talla [23] yang dilengkapi dengan pendefinisian kata tugas berdasarkan buku karangan Masnur Muslich [25] dan Abdul Muslich [26], kata tugas dipilih menjadi *stopword* karena merupakan kata yang sering digunakan namun tidak memiliki nilai informasi. Selain itu, terdapat tambahan kata-kata yang menjadi kode format dari Radio Suara Surabaya FM. Kemudian hasil dari kedua tahap ini, data diterjemahkan untuk menjadi *corpus* serta disimpan untuk penggunaan pemrosesan data selanjutnya.



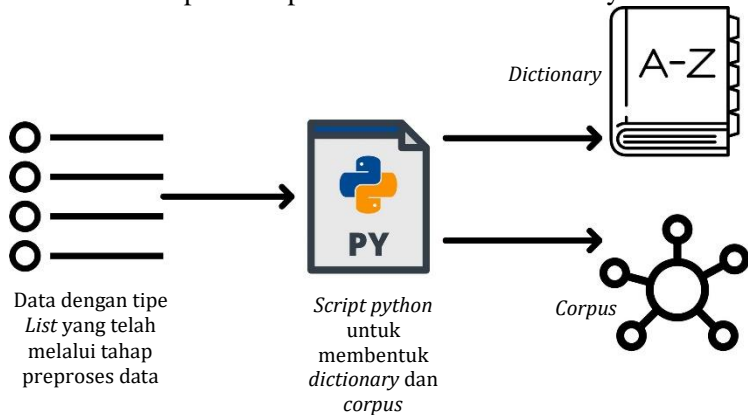
**Diagram 4.2 Alur praproses data**



### 4.3.3. *Topic modeling*

#### 4.3.3.1. Pembentukan *Dictionary* dan *Corpus*

Untuk melakukan *Topic modeling*, data perlu dirubah dalam bentuk *dictionary* dan *corpus*. *Dictionary* merupakan format data yang mengandung himpunan kata unik yang diberi indeks, sehingga dapat memudahkan dalam menampilkan kata yang termasuk dalam model. *Corpus* merupakan format data yang berbentuk dokumen *term-matrix*, digunakan dalam melakukan eksperimen pembentukan model nantinya



**Diagram 4.3** Alur pembentukan *Dictionary* dan *Corpus*

#### 4.3.3.2. Topic modeling dengan LDA

Pada tahapan proses *Topic modeling* dengan *LDA*, langkah utama yang dilakukan adalah membentuk model dengan menggunakan *library gensim*, kemudian model dievaluasi dengan evaluasi *perplexity*.

Dalam membentuk model eksperimen dilakukan pada *input parameter*. Hasil dari pencarian model akan digunakan untuk mendapatkan topik apa saja yang muncul dari analisis pada dokumen. Setelah terdapat model topik, dilakukan evaluasi *perplexity* yang sudah terdapat dalam *python* dengan modul *logging*. Model yang menunjukkan hasil *perplexity* yang lebih kecil serta tidak berubah-ubah, akan dipilih sebagai model yang akan digunakan

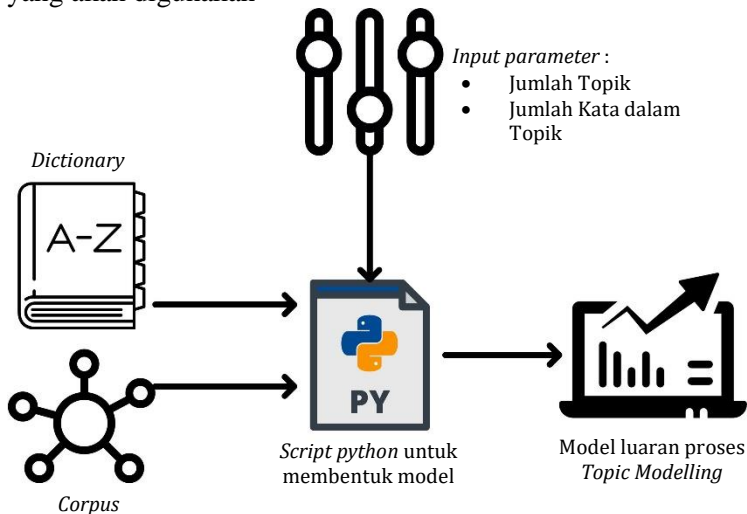


Diagram 4.4 Alur topic modeling dengan LDA

#### 4.3.4. Validasi Topik Model

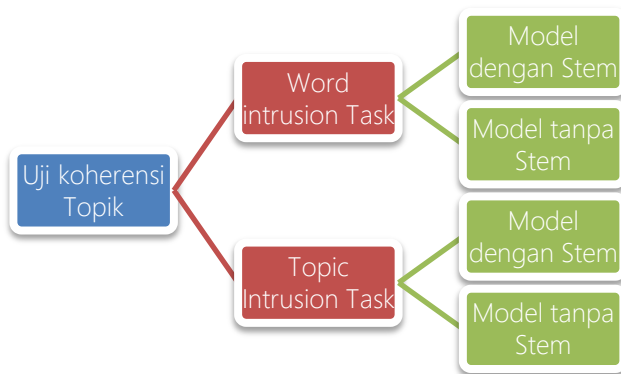
Tahap validasi topik bertujuan untuk memastikan model topik yang dihasilkan dari hasil *Topic modeling* yang dilakukan pada dokumen adalah benar, baik luaran berupa topik maupun kata-kata dalam topik.

Adapun hal hal yang dianalisis dalam tahap validasi topik model adalah

1. Jumlah iterasi yang tepat untuk membentuk topik model
2. Jumlah topik yang sesuai berdasarkan distribusi *perplexity*
3. Distribusi probabilitas masing-masing dokumen belajar terhadap model topik yang dibentuk

#### 4.3.5. Uji Koherensi Topik

Tahap uji koherensi topik dilakukan untuk menguji kemudahan topik model dalam kemampuannya untuk diinterpretasi oleh manusia. Dalam uji koherensi topik, metode yang dilakukan adalah *Topic Intrusion task* dan *Word Intrusion task* dengan membandingkan antara model yang melalui tahap *Stemming* dan model yang tidak melalui tahap *Stemming*. Penjabaran bagian-bagian kuesioner uji koherensi topik dapat dilihat pada Diagram 4.5



**Diagram 4.5 Penjabaran bagian-bagian kuesioner uji koherensi topik**

#### **4.3.5.1. Penyusunan Kuesioner *Word Intrusion Task* [13]**

*Word Intrusion task* merupakan metode uji koherensi (keterkaitan) kata-kata dalam suatu topik yang bertujuan untuk menguji keterkaitan kata dalam suatu topik dalam hal kemudahan untuk diinterpretasi oleh manusia. Metode ini dilakukan dengan menampilkan seluruh kata-kata dalam suatu topik (himpunan kata), kemudian menyisipkan kata yang tidak berhubungan diantara kata-kata tersebut. Untuk itu, peran responden dalam hal ini adalah menebak manakah satu kata yang paling tidak berhubungan diantara kata yang lainnya

##### **4.3.5.1.1. Perancangan Sistematika Kuesioner**

Tujuan dalam melakukan perancangan sistematika kuesioner adalah untuk memudahkan responden dalam memahami konteks kuesioner, sehingga dapat memberikan hasil yang lebih objektif.

Terdapat tiga bagian yaitu bagian judul, bagian deskripsi pembuka dan bagian pertanyaan. Pada bagian judul berfungsi untuk mengidentifikasi aktivitas dan metode uji koherensi topik, pada bagian deskripsi pembuka berisi pengenalan penulis, tujuan pengumpulan data, deskripsi singkat metode pengumpulan data dan pengantar cara pengisian data. Pada bagian pertanyaan, pertanyaan ditujukan agar responden memilih satu kata yang paling tidak berhubungan dari kumpulan opsi yang ditampilkan. Untuk setiap kali menampilkan kuesioner, opsi pilihan akan diacak urutannya.

##### **4.3.5.1.2. Perancangan Materi Kuesioner**

Dalam merancang materi kuesioner dengan metode *Word Intrusion Task*, hal yang perlu diperhatikan adalah kemudahan untuk dipahami oleh manusia, sehingga kuesioner dirancang dengan membagi distribusi himpunan kata ke dalam kelompok-kelompok kecil. Pengelompokan kata ke dalam kelompok kecil dilakukan dengan metode *roulette wheel*. Metode *roulette wheel* dipilih karena metode ini dapat

melakukan pemilihan kata secara acak, namun tetap memperhatikan probabilitas kata dalam topik.

Langkah langkah dalam melakukan pengelompokan kata dengan metode *roulette wheel* adalah sebagai berikut

1. Susun kata dan nilai probabilitasnya
2. Menghitung *Range* nilai sesuai probabilitas untuk setiap kata
3. Menyusun *Range* nilai untuk seluruh kata dalam deret akumulatif
4. Mendapatkan beberapa angka secara acak dengan distribusi uniform dalam rentang nilai minimal dan nilai maksimal dalam deret akumulatif
5. Menerjemahkan seluruh angka hasil pengacakan ke dalam distribusi kata sesuai deret akumulatif
6. Memindahkan kata ke dalam kelompok kecil
7. Kembali ke nomor 1 hingga jumlah kelompok kecil memenuhi jumlah yang dibutuhkan
8. Hasil pengelompokan ke kelompok kecil dengan *roulette wheel* akan menjadi opsi untuk masing-masing pertanyaan

#### **4.3.5.2. Kuesioner *Topic Intrusion Task* [13]**

*Topic Intrusion task* merupakan metode uji koherensi (keterkaitan) topik dengan dokumen/narasi yang bersangkutan. Metode ini bertujuan untuk menguji keterkaitan topik(kumpulan kata) dengan dokumen/kalimat yang dimaksud dalam hal kemudahannya untuk diinterpretasi oleh manusia. Metode ini dilakukan dengan menampilkan dokumen/narasi, dan seluruh opsi topik dalam bentuk kumpulan kata. Untuk itu, peran responden dalam hal ini adalah menebak manakah satu topik yang paling sesuai dengan dokumen/narasi diantara seluruh pilihan topik yang disediakan.

#### **4.3.5.2.1. Perancangan Sistematika Kuesioner**

Tujuan dalam melakukan perancangan sistematika kuesioner adalah untuk memudahkan responden dalam memahami konteks kuesioner, sehingga dapat memberikan hasil yang lebih objektif.

Terdapat tiga bagian yaitu bagian judul, bagian deskripsi pembuka dan bagian pertanyaan. Pada bagian judul berfungsi untuk mengidentifikasi aktivitas dan metode uji koherensi topik, pada bagian deskripsi pembuka berisi pengenalan penulis, tujuan pengumpulan data, deskripsi singkat metode pengumpulan data dan pengantar cara pengisian data. Pada bagian pertanyaan, pertanyaan ditujukan agar responden memilih satu topik (himpunan kata) yang paling tidak berhubungan dari kumpulan opsi yang ditampilkan. Untuk setiap kali menampilkan kuesioner, opsi pilihan akan diacak urutannya.

#### **4.3.5.2.1. Perancangan Materi Kuesioner *Topic Intrusion task 1***

Dalam merancang materi kuesioner dengan metode *Topic Intrusion task 1*, pilihan opsi ditampilkan secara murni hasil dari pemodelan topik dengan metode *LDA*, dalam artian seluruh tampilan himpunan kata dalam topik ditampilkan sebagai opsi pilihan yang nantinya dipilih oleh responden.

#### **4.3.5.2.2. Perancangan Materi Kuesioner *Topic Intrusion task 2***

Dalam merancang materi kuesioner dengan metode *Topic Intrusion task 2*, hal yang perlu diperhatikan adalah kemudahan untuk dipahami oleh manusia, sehingga kuesioner dirancang dengan tidak mencantumkan seluruh himpunan kata dalam topik, melainkan melakukan pemilihan secara acak sesuai probabilitas kata dalam topik kedalam kelompok dengan jumlah yang lebih sedikit. Pemilihan kata secara acak ke dalam kelompok kecil dilakukan dengan metode *roulette wheel*.

Metode *roulette wheel* dipilih karena metode ini dapat melakukan pemilihan kata secara acak, namun tetap memperhatikan probabilitas kata dalam topik.

Langkah langkah dalam melakukan pengelompokan kata dengan metode *roulette wheel* adalah sebagai berikut

1. Susun kata dan nilai probabilitasnya
2. Menghitung *Range* nilai sesuai probabilitas untuk setiap kata
3. Menyusun *Range* nilai untuk seluruh kata dalam deret akumulatif
4. Mendapatkan beberapa angka secara acak dengan distribusi uniform dalam rentang nilai minimal dan nilai maksimal dalam deret akumulatif
5. Menerjemahkan seluruh angka hasil pengacakan ke dalam distribusi kata sesuai deret akumulatif
6. Memindahkan kata ke dalam kelompok kecil
7. Kembali ke nomor 1 hingga jumlah kelompok kecil memenuhi jumlah yang dibutuhkan
8. Hasil pengelompokan ke kelompok kecil dengan *roulette wheel* akan menjadi opsi untuk masing-masing pertanyaan

#### **4.3.5.3. Analisis Hasil Kuesioner**

Setelah kuesioner dikumpulkan, tahap selanjutnya adalah analisis hasil kuesioner. Analisis hasil kuesioner dilakukan dengan menghitung jawaban yang tepat dari berbagai pilihan yang disediakan. Hasil jawaban responden kemudian dikalkulasi dengan satuan persen

*Halaman ini sengaja dikosongkan*



## BAB V IMPLEMENTASI

Pada bab ini berisi tentang proses implementasi dalam pembuatan sistem. Proses implementasi dilakukan dengan menggunakan Bahasa *Python*.

### 5.1. Perangkat Penelitian

Dalam pelaksanaan analisis topik informasi publik media sosial di surabaya menggunakan pemodelan *Latent Dirichlet Allocation (LDA)*, dibutuhkan perangkat-perangkat untuk menunjang keberlangsungan penelitian. Adapun perangkat-perangkat yang dibutuhkan berupa perangkat keras dan perangkat lunak, dengan spesifikasi sebagai berikut :

#### 1. Perangkat keras

*Processor* : *Intel (R) Core (TM) i5-4200 CPU  
@ 1.60GHz (4 CPUs)*  
*Memory* : *8192 MB RAM*

#### 2. Perangkat lunak

*Operating SyStem* : *Windows10 Pro 64-bit*  
*Bahasa Pemrograman* : *Python 3.5 64-bit*  
*Software Tools* : *IDE Pycharm 2016.3.2*  
*DBMS MySQL*  
*Ms. Excel*  
*Minitab*  
*Library* : *Anaconda 4.2.0*  
*Gensim*

## 5.2. Mempersiapkan Data

Mempersiapkan data merupakan tahapan yang dilakukan untuk mengolah data agar siap untuk digunakan. Ada 2 tahapan utama dalam melakukan implementasi penelitian yaitu memuat data dan membersihkan data.

### 5.2.1. Memuat Data

Sesuai dengan hasil justifikasi data yang dilakukan pada tahap observasi data mentah, bahwa data yang digunakan adalah pesan yang dinilai memiliki kandungan informasi yang baik, yaitu pesan berita yang dikirim melalui akun *Facebook* E100 dan *tweet* dari masyarakat yang di-*tweet* kembali oleh akun e100ss. Kemudian, kedua sumber data tersebut yang tersimpan dalam format *.csv* dimuat dalam variabel '*tweet*' untuk data *tweet* dari masyarakat yang di-*tweet* kembali oleh akun e100ss dan variabel '*fbpost*' untuk data pesan berita yang dikirim melalui akun *Facebook* E100. Setelah kedua sumber data disimpan dalam dua variabel yang berbeda, kedua data disatukan dalam bentuk *list* dengan nama '*listAgregat*'. *Source code* yang diimplementasikan untuk memuat data ditampilkan pada Kode 5.1

```
tweet = open('input/retweet-oleh-  
e100ss.csv')  
fbPost = open('input/post-fb-e100ss.csv')  
  
listAgregat = []  
for i in tweet:  
    listAgregat.append(i)  
for j in fbPost:  
    listAgregat.append(j)
```

**Kode 5.1 Memuat (Loading) Data**

### 5.2.2. Membersihkan Data

Pada tahap membersihkan data, *library* yang digunakan adalah *library* “re” dan *library* “preprocessor”. Tahap membersihkan data diawali dengan membuat *list* sebagai wadah untuk menyimpan data yang sudah dibersihkan dengan nama ‘list\_data\_bersih’. Kemudian untuk melakukan pembersihan data, dilakukan secara bertahap dengan rincian sebagai berikut: dataBersih1 membersihkan *url*, *mention*, *reserved words* dan *number*, dataBersih2 membersihkan karakter *non-alphanumeric*, dataBersih3 membersihkan token sepanjang 1 digit. Setelah data dibersihkan, data disimpan dalam *list* ‘list\_data\_bersih’. *Source code* yang diimplementasikan untuk membersihkan data ditampilkan pada Kode 5.2

```
import re
pattern = re.compile('\W')
import preprocessor as p

list_data_bersih = []

for k in listAgregat:
    p.set_options(p.OPT.URL, p.OPT.MENTION,
p.OPT.RESERVED, p.OPT.NUMBER)
    dataBersih1 = p.clean(k)
    dataBersih2 = re.sub(pattern, ' ',
dataBersih1)
    dataBersih3 = re.sub(r'\b\w{1,1}\b', '',
dataBersih2)
    list_data_bersih.append(dataBersih3)
```

**Kode 5.2 Membersihkan data**

### 5.3. Pra-Proses Data

Tahap pra-proses data merupakan tahapan yang dilakukan sebelum pemrosesan data agar tahap pemrosesan data dapat berjalan dengan optimal. Ada 4 tahapan utama dalam melakukan implementasi penelitian yaitu pendefinisian *stopwords*, menghapus *stopwords* dari dokumen, mengkonversi

dokumen ke dalam bentuk *dictionary* dan mengkonversi *dictionary* ke dalam bentuk matriks dokumen atau *corpus*.

### 5.3.1. Pendefinisian *Stopwords*

Dalam mendefinisikan *stopwords*, *stopwords* disimpan dalam *file* berformat *.csv* yang terdiri dari satu kata per baris, untuk kemudian dimuat dalam *list* dengan nama ‘*list\_stopword*’ yang mana *list* tersebut telah dibuat sebelumnya. *Source code* yang diimplementasikan untuk memuat *stopword* ditampilkan pada Kode 5.3

```
list_stopword = []
with open('file/StopwordKataTugas.csv', 'r')
as stopwords:
    for line in stopwords:
        list_stopword.append(line.strip())
```

**Kode 5.3** Pendefinisian *stopwords*

### 5.3.2. Menghapus *Stopwords* dari Dokumen

Setelah *stopwords* didefinisikan dan disimpan dalam *list* dan dokumen telah tersimpan dalam *list*, langkah selanjutnya adalah dengan menghapus *stopwords* yang terdapat dalam dokumen. Dalam menghapus *stopwords* dari dokumen, perlu dilakukan tokenisasi. Adapun *library* yang digunakan adalah ‘*nlTK.tokenize*’ dengan modul ‘*RegexpTokenizer*’. Adapun tahap-tahap yang dilakukan adalah untuk setiap dokumen dalam *corpus* yang tersimpan dalam ‘*list\_data\_bersih*’ diubah menjadi *lowercase* dengan tipe data *string*, kemudian dokumen dilakukan tokenisasi sehingga dokumen dipecah perkata, kata-kata yang hanya terdiri dari angka dihapus, kemudian untuk setiap token yang tercatat sebagai *stopword* dalam ‘*list\_stopword*’ dihapus dari dokumen, pada akhirnya data yang telah bersih dari *stopword* disimpan dalam *list* ‘*tokenlist\_dataAgregat*’ yang telah dibuat sebelumnya. *Source code* yang diimplementasikan untuk menghapus *stopword* dari dokumen ditampilkan pada Kode 5.4

```

from nltk.tokenize import RegexpTokenizer
tokenizer = RegexpTokenizer(r'\w+')

tokenlist_dataAgregat= []
for i in dataPreparation.list_data_bersih:
    dataKasar0 = str(i.lower())
    tokens = tokenizer.tokenize(dataKasar0)
    tokenWithoutInt = [item for item in tokens
if not item.isdigit()]
    stopped_tokens = [i for i in
tokenWithoutInt if not i in list_stopword]

tokenlist_dataAgregat.append(stopped_tokens)

```

**Kode 5.4 Menghapus *stopwords* dari dokumen**

### 5.3.3. Mengkonversi Dokumen ke Dalam bentuk *Dictionary*

Tahap berikutnya adalah tahap mengkonversi dokumen ke dalam bentuk *dictionary*, tujuan dari tahap ini adalah menyimpan data token ke dalam bentuk *dictionary*, kemudian memberikan indeks yang berfungsi untuk mengidentifikasi kata tersebut. Dalam melakukan konversi dokumen ke dalam bentuk *dictionary* digunakan *library gensim* dengan modul *corpora*. *File dictionary* kemudian disimpan dengan nama 'dictionary\_data\_agregat.dict' agar lebih mudah untuk digunakan dalam melakukan eksperimen pada tahap-tahap berikutnya. *Source code* yang diimplementasikan untuk mengkonversi dokumen ke dalam bentuk *dictionary* ditampilkan pada Kode 5.5

```

# turn our tokenized documents into a id <->
term dictionary
from gensim import corpora

dictionary_dataAgregat =
corpora.Dictionary(tokenlist_dataAgregat)
dictionary_dataAgregat.save('file/dictionary
_dataAgregat.dict')

```

### Kode 5.5 Mengkonversi dokumen ke dalam bentuk *dictionary*

#### 5.3.4. Mengkonversi *Dictionary* dan *List* ke dalam Matriks Dokumen

Setelah terdapat *file dictionary* hasil konversi dokumen, tahap selanjutnya adalah mengkonversi *dictionary* dalam matriks dokumen yang selanjutnya disebut *corpus*. Adapun *library* yang digunakan adalah *gensim* dengan modul *corpora*. *Corpus* tersebut kemudian disimpan dalam *file* dengan nama 'corpora.mm' agar lebih mudah untuk digunakan dalam melakukan eksperimen pada tahap-tahap berikutnya. *Source code* yang diimplementasikan untuk mengkonversi *dictionary* dan *list* ke dalam matriks dokumen ditampilkan pada Kode 5.6

```
# convert tokenized documents into a
document-term matrix
from gensim import corpora

corpus = [dictionary_dataAgregat.doc2bow(i)
for i in tokenlist_dataAgregat]
print("corpus")
print(corpus)
corpora.MmCorpus.serialize('file/corpora.mm'
, corpus)
```

### Kode 5.6 Mengkonversi *dictionary* dan *list* ke dalam matriks dokumen

#### 5.4. Pemodelan Topik dengan *Latent Dirichlet Allocation*

Tahapan pemodelan topik dengan *Latent Dirichlet Allocation* merupakan tahapan yang dilakukan untuk membentuk model topik. Terdapat dua hal penting yang menjadi perhatian dalam tahap ini, yaitu alur pemodelan topik dengan *Latent Dirichlet Allocation* dan eksperimen pemodelan topik

### 5.4.1. Alur Pemodelan Topik dengan *Latent Dirichlet Allocation*

#### 5.4.1.1. Loading Dictionary dan Corpus

Dalam konteks ini, *loading* data yang dimaksud adalah memuat data *dictionary* dan *corpus* yang pada tahap sebelumnya telah disimpan *file*. Agar lebih mudah dalam pengolahan, data disimpan dalam variabel, variabel ‘*dictionary\_dataAgregat*’ untuk *file* ‘*dictionary\_dataAgregat.dict*’ dan variabel ‘*corpus*’ untuk *file* ‘*file/corpora.mm*’. *Source code* yang diimplementasikan untuk *loading dictionary* dan *Corpus* ditampilkan pada Kode 5.7

```
dictionary_dataAgregat =
Dictionary.load("file/dictionary_dataAgregat
.dict")
corpus = MmCorpus("file/corpora.mm")
```

**Kode 5.7 Loading Dictionary dan Corpus**

#### 5.4.1.2. Pembentukan Model Topik

Pada tahap pembentukan model topik *library* yang digunakan adalah *gensim* dengan modul *MmCorpus* dan *Dictionary*. Dalam pembentukan model topik, diperlukan *input parameter*, yaitu jumlah topik, jumlah kata dalam topik, dan *passes*. Adapun yang dimaksud dengan *passes* adalah jumlah iterasi dalam pembentukan model topik. Ketiga *input parameter* ini nantinya akan dilakukan eksperimen untuk mencari nilai *perplexity* yang optimal. Nilai *perplexity* yang semakin kecil menunjukkan model yang dibentuk semakin baik. *Source code* yang diimplementasikan untuk eksperimen pada *input parameter LDA* ditampilkan pada Kode 5.8

```
import gensim
from gensim.corpora import MmCorpus,
Dictionary

jumlahTopik = 15
jumlahKata = 10
```

```
# generate LDA model
LDAModel =
gensim.models.LDAModel.LDAModel(corpus,
num_topiks=jumlahTopik,
id2word=dictionary_dataAgregat, passes=50)
print(LDAModel.print_topiks(num_topiks=jumla
hTopik, num words=jumlahKata))
```

**Kode 5.8 Eksperimen pada input parameter LDA**

#### 5.4.1.3. Pendokumentasian *Logging*

Dalam melakukan eksperimen, diperlukan *logging* untuk mengetahui catatan terkait kejadian yang terjadi dalam proses pembentukan model topik. Catatan penting yang dibutuhkan adalah nilai *perplexity* yang mana nilai *perplexity* sudah terkalkulasi secara otomatis sebagai fitur modul *gensim*. Untuk melakukan *logging*, diperlukan *library logging*, kemudian *file logging* disimpan dalam *file* berformat *.csv* dengan nama sesuai kebutuhan. Dalam contoh kasus di bawah, nama *file* disesuaikan dengan jumlah topik dan jumlah kata. *Source code* yang diimplementasikan untuk pendokumentasian *logging* ditampilkan pada Kode 5.9

```
import logging

namaFile = str(jumlahTopik) + '~'
+str(jumlahKata)
namaFileLog = 'logs/' + namaFile + '~' +
'.csv'

info =
logging.basicConfig(filename=namaFileLog,
filemode='w', format='%(asctime)s :
%(levelname)s : %(message)s',
level=logging.INFO)
```

**Kode 5.9 Pendokumentasian *Logging***



### 5.4.2. Eksperimen Pemodelan Topik dengan LDA

Tahapan eksperimen pemodelan topik dengan *Latent Dirichlet Allocation* merupakan tahapan yang dilakukan untuk membentuk model topik terbaik dengan melakukan eksperimen pada *input parameter*, yaitu *passes* atau jumlah iterasi dan jumlah topik. Sehingga eksperimen akan dibedakan kedalam dua tahap, yaitu penentuan jumlah iterasi dan penentuan jumlah topik.

#### 5.4.2.1. Penentuan Jumlah Iterasi

Dalam LDA, istilah iterasi dikenal dengan *passes*. Penentuan iterasi merupakan tahap yang penting dalam menentukan model, hal ini untuk menghasilkan model yang terbaik, dalam konteks ini apabila jumlah iterasi terlalu sedikit, akan menghasilkan model yang belum stabil dan *under fitting*, sementara iterasi yang terlalu banyak akan menghasilkan model yang *overfitting*. Penentuan jumlah *passes* dilakukan dengan melakukan eksperimen pada jumlah topik. Penentuan jumlah *passes* diawali dengan memberikan nilai mula-mula sebesar 50, kemudian jumlah topik ditentukan yaitu sebanyak 5 kali, yaitu 10, 20, 30, 40, 50 topik. Berdasarkan eksperimen jumlah topik, nilai *perplexity* yang muncul akan dicatat untuk dianalisis tren nilainya secara visual dan dilakukan penghitungan standar deviasinya. Sehingga pada akhirnya nilai *passes* yang akan digunakan adalah nilai *passes* yang paling awal setelah menunjukkan tren yang stabil. *Source code* yang diimplementasikan untuk penentuan jumlah iterasi ditampilkan pada Kode 5.10

```
import gensim
from gensim.corpora import MmCorpus,
Dictionary

jumlahTopik = 15
jumlahKata = 10

# generate LDA model
```

```
LDAModel =
gensim.models.LDAModel.LDAModel(corpus,
num_topiks=jumlahTopik,
id2word=dictionary_dataAgregat, passes=50)
print(LDAModel.print_topiks(num_topiks=jumla
hTopik, num_words=jumlahKata))
```

#### Kode 5.10 Penentuan jumlah iterasi

##### 5.4.2.2. Penentuan Jumlah Topik

Setelah menentukan jumlah iterasi, eksperimen dilakukan pada jumlah topik. Eksperimen jumlah topik merupakan tahap yang penting dalam menentukan model, hal ini untuk menghasilkan model yang terbaik, dalam konteks ini model yang baik adalah model dengan nilai *perplexity* yang rendah, semakin rendah nilai *perplexity*, menunjukkan akurasi model lebih baik. Penentuan jumlah topik dilakukan dengan melakukan eksperimen pada nilai jumlah topik. Penentuan jumlah topik diawali dengan memberikan nilai mula-mula yaitu sebanyak 5 kali, yaitu 10, 20, 30, 40, 50 topik. Berdasarkan eksperimen jumlah topik, nilai *perplexity* yang muncul akan dicatat untuk dianalisis tren nilainya secara visual dan dilakukan penghitungan standar deviasinya. Sehingga pada akhirnya jumlah topik yang dipilih adalah jumlah topik yang memiliki nilai rata-rata paling rendah dengan standar deviasi minimum. *Source code* yang diimplementasikan untuk penentuan jumlah topik ditampilkan pada Kode 5.11

```
import gensim
from gensim.corpora import MmCorpus,
Dictionary

jumlahTopik = 15
jumlahKata = 10

# generate LDA model
LDAModel =
gensim.models.LDAModel.LDAModel(corpus,
```

```
num_topiks=jumlahTopik,
id2word=dictionary_dataAgregat, passes=20)
```

#### Kode 5.11 Penentuan jumlah topik

### 5.4.2.3. Menyimpan Model

Model dengan jumlah iterasi dan jumlah topik yang telah ditentukan perlu disimpan untuk dapat digunakan kembali dengan cepat. Model disimpan dalam format *.model*. Adapun cara menyimpan model ditampilkan pada Kode 5.12

```
namaFile = str(jumlahTopik) + '~'
+str(jumlahKata)
namaFileModel = 'models/' + namaFile +
'.model'

LDAModel.save(namaFileModel)
```

#### Kode 5.12 Menyimpan model

Dalam tugas akhir ini, ‘namaFileModel’ yang digunakan adalah jumlah topik dan jumlah kata yang digunakan untuk eksperimen pembentukan model

## 5.5. Validasi Model Topik

### 5.5.1. Klasifikasi Dokumen ke dalam Topik dengan Metode LDA

Tahap klasifikasi dokumen ke dalam topik dengan metode *LDA* bertujuan untuk mengklasifikasikan setiap dokumen ke dalam topik yang telah dibentuk sebelumnya. Secara umum, tahapan dalam melakukan klasifikasi adalah melakukan import *library*, *loading input*, operasi data, dan menampilkan distribusi dokumen dalam topik.

#### 5.5.1.1. Import Library

Untuk melakukan klasifikasi data, diperlukan modul *MmCorpus* dan *Dictionary* yang terdapat pada *library gensim*,

maka dari itu perlu dilakukan import *library* dengan notasi yang dapat dilihat pada Kode 5.13

```
import gensim
from gensim.corpora import MmCorpus,
Dictionary
```

**Kode 5.13 Import library**

#### 5.5.1.2. Loading Input

Terdapat empat *input* yang digunakan untuk melakukan klasifikasi data, yaitu model, *corpus*, *dictionary* dan *input* dokumen. *Input* model yang dipilih merupakan model (dengan format *.model*) yang sudah dibentuk sebelumnya dan terpilih sebagai model untuk melakukan klasifikasi. *Corpus* dan *dictionary* dijadikan *input* merupakan *corpus* dan *dictionary* yang telah dibentuk pada tahap prapemrosesan data. *Input* dokumen merupakan *file* yang berisi kumpulan dokumen yang akan diklasifikasi. *Source code* yang diimplementasikan untuk *loading input* ditampilkan pada Kode 5.14

```
# Load model.
filename = 'models/4~10.model'
model =
gensim.models.LDAModel.load(filename)

#Loading Dictionary dan Corpus
dictionary_dataAgregat =
Dictionary.load("file/dictionary_dataAgregat
.dict")
corpus = MmCorpus("file/corpora.mm")

#Loading input dokumen
tweet = open('input/retweet-oleh-
e100ss.csv')
fbPost = open('input/post-fb-e100ss.csv')
```

**Kode 5.14 Loading input**

### 5.5.1.3. Operasi Data

Tahap operasi data dibagi menjadi tiga tahap, mengubah kumpulan dokumen ke dalam *list* dokumen, pembentukan *list* data hasil klasifikasi, dan klasifikasi data. Tahap mengubah kumpulan dokumen ke dalam *list* dokumen bertujuan untuk menyatukan berbagai sumber dokumen yang nantinya akan diklasifikasikan dalam bentuk *list* agar mudah dalam melakukan klasifikasi. Tahap pembentukan *list* data hasil klasifikasi bertujuan untuk membentuk *list* sebagai wadah yang nantinya akan diisi oleh dokumen-dokumen yang sudah diklasifikasi, *list* data dibuat sejumlah topik yang digunakan untuk mengklasifikasi. Tahap klasifikasi data bertujuan untuk melakukan klasifikasi data ke dalam masing-masing topik dengan mengabaikan dokumen yang tidak condong ke topik manapun, sehingga untuk setiap *list* akan tersimpan dokumen dari masing-masing dokumen yang terklasifikasi dalam topik tersebut. *Source code* yang diimplementasikan untuk operasi data ditampilkan pada Kode 5.15

```
#Mengubah kumpulan dokumen ke dalam list
dokumen
listAgregat = []
for i in tweet:
    listAgregat.append(i)
for j in fbPost:
    listAgregat.append(j)

#Pembentukan list data hasil klasifikasi
k = 4;
listDocument = [[] for z in Range(k)]

#Klasifikasi data
listTopikNo = []
for document in listAgregat:
    document1 = document.split()
    document2 =
dictionary_dataAgregat.doc2bow(document1)
print model[document2]
```

```

a = list(sorted(model[document2],
key=lambda x: x[1]))
print a[-1]
topikNo, probability = a[-1]
listDocument[topikNo].append(document)
listTopikNo.append(topikNo)
print document

```

#### Kode 5.15 Operasi data

#### 5.5.1.4. Menampilkan Distribusi Dokumen dalam Topik

Setelah dokumen diklasifikasikan, untuk mempermudah dalam menganalisis distribusi dokumen dalam topik, distribusi dokumen dalam topik perlu ditampilkan dengan notasi yang dapat dilihat pada Kode 5.16

```

b = sorted(listTopikNo)
print b[-1]

for topikNo in Range(b[0], b[-1]+1):
    judul = "Topik #" + str(topikNo)
    print judul + " : " +
    str(listDocument[topikNo])

```

#### Kode 5.16 menampilkan distribusi dokumen dalam topik

### 5.5.2. Analisis Distribusi Probabilitas Dokumen Per Topik

#### 5.5.2.1. Analisis Distribusi Probabilitas Dokumen Per Topik dengan Metode *LDA*

##### 5.5.2.1.1. Import *Library*

Untuk melakukan klasifikasi data, diperlukan modul *MmCorpus* dan *Dictionary* yang terdapat pada *library gensim*, maka dari itu perlu dilakukan import *library* dengan notasi yang dapat dilihat pada Kode 5.17

```
import gensim
from gensim.corpora import MmCorpus,
Dictionary
```

#### Kode 5.17 Import library

##### 5.5.2.1.2. Loading Input

Terdapat empat *input* yang digunakan untuk melakukan klasifikasi data, yaitu *model*, *corpus*, *dictionary* dan *input* dokumen. *Input* model yang dipilih merupakan model (dengan format *.model*) yang sudah dibentuk sebelumnya dan terpilih sebagai model untuk melakukan klasifikasi. *Corpus* dan *dictionary* dijadikan *input* merupakan *corpus* dan *dictionary* yang telah dibentuk pada tahap prapemrosesan data. *Input* dokumen merupakan *file* yang berisi kumpulan dokumen yang akan diklasifikasi. *Source code* yang diimplementasikan untuk *loading input* ditampilkan pada Kode 5.18

```
# Load model.
filename = 'models/4~10.model'
model =
gensim.models.LDAModel.load(filename)

# Print model
for i in Range(0, model.num_topiks):
    print model.print_topik(i)

# Load Input Dictionary dan Corpus
Dictionary.load("file/dictionary_dataAgregat
.dict")
corpus = MmCorpus("file/corpora.mm")
```

#### Kode 5.18 Loading input

##### 5.5.2.1.3. Operasi Data

Tahap operasi data dibagi menjadi dua tahap, inisiasi *list* data, dan klasifikasi data. Tahap inisiasi *list* data bertujuan untuk membentuk *list* sebagai wadah yang nantinya akan diisi

oleh dokumen-dokumen yang sudah diklasifikasi, *list* data dibuat sejumlah topik yang digunakan untuk mengklasifikasi. Tahap klasifikasi data bertujuan untuk melakukan klasifikasi data ke dalam masing-masing topik dengan mengabaikan dokumen yang tidak condong ke topik manapun, sehingga untuk setiap *list* akan tersimpan probabilitas dari masing-masing dokumen yang terklasifikasi dalam topik tersebut. *Source code* yang diimplementasikan untuk operasi data ditampilkan pada Kode 5.19



```

# Inisiasi List
i = 200
countTopik = [0 for x in Range(i)]
j = 200;
docsPerTopik = [[] for y in Range(j)]
countTopikKosong = 0
listTopikNo = []
k = 200;
listProbability = [[] for z in Range(k)]

# Klasifikasi Data
for doc in Range(0, corpus.num_docs):
#corpus.num_docs
    print model[corpus[doc]]
    a = list(sorted(model[corpus[doc]],
key=lambda x: x[1]))

    if len(a) == 0:
        print "kosong"
        countTopikKosong += 1
    else:
        if len(a) > 0:
            print (a[-1])
            topikNo, probability = a[-1]
            topikNoMax, probabilityMax = a[-1]
            topikNoMin, probabilityMin = a[0]
            if len(a) > 1 and probabilityMin ==
probabilityMax:
                countProbabilityMerata += 1
            else:
                countTopik[topikNo] += 1
                listTopikNo.append(topikNo)

docsPerTopik[topikNo].append([corpus[doc]])

listProbability[topikNo].append(probability)

```

**Kode 5.19 Operasi data**

#### 5.5.2.1.4. Menampilkan Jumlah Distribusi Dokumen Per Topik

Setelah dokumen diklasifikasikan, untuk mempermudah dalam menganalisis distribusi probabilitas dari dokumen dalam topik, distribusi probabilitas dari dokumen dalam topik perlu ditampilkan dengan notasi yang ditampilkan pada Kode 5.20

```
b = sorted(listTopikNo)
print b[-1]

for topikNo in Range(0, b[-1]+1): #b[-1]+1
    judul = "Topik #" + str(topikNo)
    print judul + " : " +
    str(listProbability[topikNo])
```

**Kode 5.20 Menampilkan jumlah distribusi dokumen per topik**

### 5.5.3. Visualisasi Distribusi Probabilitas Dominan dari Dokumen per Topik

Sebagai bagian dari tahap validasi topik model, distribusi probabilitas dominan dari dokumen per topik perlu divisualisasikan dengan tujuan untuk memudahkan dalam menyimpulkan apakah distribusi probabilitas cukup memberikan hasil yang meyakinkan melalui kecenderungan distribusinya.

#### 5.5.3.1. Distribusi Probabilitas Dominan dari Dokumen per Topik dengan Metode LDA

##### 5.5.3.1.1. Source code

Untuk menampilkan distribusi probabilitas dominan dari dokumen per topik dalam bentuk *list* dilakukan dengan *Source code* yang dapat dilihat pada Kode 5.21.

```
for topikNo in Range(0, b[-1]+1):
    judul = "Topik #" + str(topikNo)
```

```
print judul + " : " +
str(listProbability[topikNo])
```

**Kode 5.21** *Source code* distribusi probabilitas dominan dari dokumen per topik dengan metode *LDA*

### 5.5.3.1.1. Histogram

*Source code* diatas akan menampilkan distribusi probabilitas dominan dari dokumen per topik dalam bentuk *list*, sehingga masih sulit untuk dilakukan analisis, sehingga akan lebih mudah untuk melakukan analisis apabila divisualisasikan dalam bentuk histogram. Untuk memvisualisasikan ke dalam histogram, dibutuhkan *library* 'matplotlib.pyplot'. *Source code* yang diimplementasikan untuk eksperimen pada *input parameter LDA* ditampilkan pada Kode 5.22

```
import matplotlib.pyplot as plt
for topikNo in Range(0, b[-1]+1): #b[-1]+1
    gaussian_numbers =
listProbability[topikNo]
    judul = "Topik #" + str(topikNo)
    plt.hist(gaussian_numbers, bins=50)
    plt.title(judul)
    plt.xlabel("Probability Value")
    plt.ylabel("Frequency")
    plt.show()
```

**Kode 5.22** Histogram distribusi probabilitas dominan dari dokumen per topik dengan metode *LDA*

## 5.6. Uji Koherensi Topik

Analisis topik merupakan tahap untuk menganalisis topik yang menjadi luaran dari tahap pembentukan model topik. Analisis topik dilakukan dengan melakukan pengamatan terhadap seluruh distribusi topik termasuk distribusi kata dalam topik

### 5.6.1. Menampilkan Topik

Sebagai langkah awal untuk menganalisis topik, daftar topik perlu ditampilkan dengan Kode 5.23

```
print(LDAmode1.print_topics(-1))
```

#### Kode 5.23 Menampilkan topik

Fungsi `print_topics(-1)` merupakan sebuah fungsi yang digunakan untuk mengurutkan probabilitas distribusi kata yang terdapat dalam topik sehingga lebih mudah untuk diamati

### 5.6.2. Penyusunan Materi Kuesioner

Penyusunan materi kuesioner dibedakan menjadi empat, yaitu *Word Intrusion task* dengan *Stem*, *Word Intrusion task* tanpa *Stem*, *Topic Intrusion task* dengan *Stem* dan *Word Intrusion task* tanpa *Stem*

#### 5.6.2.1. *Word Intrusion task*

Sesuai dengan langkah-langkah penyusunan materi kuesioner *Word Intrusion Task*, implementasi materi kuesioner dilakukan dengan menyusun tabel deret akumulatif untuk masing-masing topik, kemudian dilanjutkan dengan melakukan pemilihan kata dalam topik secara acak dengan metode *roulette wheel* sehingga menghasilkan kelompok-kelompok kecil yang nantinya akan diuji keterkaitan setiap katanya, pada akhirnya, untuk setiap kelompok kecil, disisipkan satu kata lain yang memiliki probabilitas kecil terhadap topik tersebut. Berikut merupakan deret akumulatif dan hasil pengacakan dengan *roulette wheel* untuk setiap topik yang ditampilkan pada Tabel 5.1 sampai Tabel 5.16

##### 5.6.2.1.1. *Word Intrusion task* dengan *Stem*

**Tabel 5.1 Deret Akumulatif Topik #0**

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
surabaya	0.021	21	21	0-21
kawan	0.01	10	31	22-31
indonesia	0.008	8	39	32-39
suara	0.007	7	46	40-46
selamat	0.006	6	52	47-52
news	0.005	5	57	53-57
anak	0.005	5	62	58-62

wib	0.005	5	67	63-67
foto	0.005	5	72	68-72
orang	0.005	5	77	73-77
radio	0.004	4	81	78-81
ibadah	0.004	4	85	82-85
ciri	0.004	4	89	86-89
hilang	0.004	4	93	90-93
kumandang	0.004	4	97	94-97
adzan	0.004	4	101	98-101
tunai	0.004	4	105	102-105
jatim	0.004	4	109	106-109
tiket	0.003	3	112	110-112
rumah	0.003	3	115	113-115

**Tabel 5.2 Pengacakan dengan *Roulette wheel* pada Topik #0**

Pengacakan ke-	Angka yang muncul	Kata
1	51	selamat
	74	orang
	83	ibadah
	100	adzan
	115	rumah
2	88	ciri
	90	hilang
	74	orang
	113	rumah
	58	anak
3	106	tiket
	105	tunai
	78	radio
	107	jatim
	4	surabaya
4	32	indonesia
	15	surabaya

	108	jatim
	77	orang
	71	foto
5	94	kumandang
	43	selamat
	104	tunai
	66	wib
	84	ibadah

Tabel 5.3 Deret Akumulatif Topik #1

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
hujan	0.017	17	17	0-17
surabaya	0.016	16	33	18-33
pagi	0.015	15	48	34-48
kawan	0.014	14	62	49-62
jalan	0.014	14	76	63-76
jl	0.011	11	87	77-87
banjir	0.011	11	98	88-98
selamat	0.01	10	108	99-108
kawasan	0.01	10	118	109-118
air	0.009	9	127	119-127
cuaca	0.009	9	136	128-136
wib	0.008	8	144	137-144
kota	0.007	7	151	145-151
deras	0.006	6	157	152-157
daerah	0.006	6	163	158-163
gubernur	0.006	6	169	164-169
hati	0.005	5	174	170-174
listrik	0.005	5	179	175-179
kerja	0.005	5	184	180-184

suhu	0.005	5	189	185-189
------	-------	---	-----	---------

**Tabel 5.4 Pengacakan dengan *Roulette wheel* pada Topik #1**

Pengacakan ke-	Angka yang muncul	Kata
1	32	surabaya
	149	kota
	161	daerah
	181	kerja
	167	gubernur
2	177	<i>listrik</i>
	124	air
	146	kota
	23	surabaya
	111	kawasan
3	152	deras
	122	air
	95	banjir
	114	kawasan
	46	pagi
4	2	hujan
	134	cuaca
	39	pagi
	143	wib
	187	suhu
5	135	cuaca
	19	surabaya
	66	jalan
	155	deras
	93	banjir

**Tabel 5.5 Deret Akumulatif Topik #2**

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
-----------------	--------------	---------------------	------------------	-----------------------------



info	0.035	35	35	0-35
celaka	0.023	23	58	36-58
lokasi	0.019	19	77	59-77
motor	0.015	15	92	78-92
truk	0.015	15	107	93-107
foto	0.015	15	122	108-122
kendara	0.015	15	137	123-137
via	0.014	14	151	138-151
mobil	0.014	14	165	152-165
tugas	0.013	13	178	166-178
lapor	0.013	13	191	179-191
korban	0.012	12	203	192-203
libat	0.012	12	215	204-215
jalan	0.01	10	225	216-225
bakar	0.01	10	235	226-235
update	0.01	10	245	236-245
sepeda	0.01	10	255	246-255
data	0.009	9	264	256-264
kronologi	0.009	9	273	265-273
dengar	0.008	8	281	274-281

**Tabel 5.6 Pengacakan dengan *Roulette wheel* pada Topik #2**

Pengacakan ke-	Angka yang muncul	Kata
1	235	bakar
	203	korban
	84	motor
	271	kronologi
	246	sepeda
2	242	update
	117	foto

	142	via
	268	kronologi
	34	info
3	238	update
	256	data
	8	info
	191	lapor
	55	celaka
4	279	dengar
	184	lapor
	63	lokasi
	217	jalan
	199	korban
5	219	jalan
	99	truk
	148	via
	112	foto
	65	lokasi

**Tabel 5.7 Deret Akumulatif Topik #3**

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
arah	0.073	73	73	0-73
padat	0.069	69	142	74-142
macet	0.044	44	186	143-186
jalur	0.03	30	216	187-216
tol	0.027	27	243	217-243
imbas	0.024	24	267	244-267
via	0.021	21	288	268-288
foto	0.02	20	308	289-308
truk	0.017	17	325	309-325
waru	0.016	16	341	326-341

lintas	0.012	12	353	342-353
simpang	0.012	12	365	354-365
surabaya	0.012	12	377	366-377
jalan	0.012	12	389	378-389
lajur	0.012	12	401	390-401
mogok	0.01	10	411	402-411
waspada	0.01	10	421	412-421
arus	0.01	10	431	422-431
raya	0.009	9	440	432-440
sidoarjo	0.009	9	449	441-449

**Tabel 5.8 Pengacakan dengan *Roulette wheel* pada Topik #3**

Pengacakan ke-	Angka yang muncul	Kata
1	172	imbis
	187	jalur
	274	via
	32	arah
	110	padat
2	443	raya
	354	simpang
	344	lintas
	6	arah
	125	padat
3	331	waru
	209	jalur
	147	macet
	138	padat
	375	surabaya
4	241	tol
	104	padat
	172	macet
	250	imbis
	315	truk

5	412	mogok
	425	waspada
	335	truk
	439	arus
	256	tol

### 5.6.2.1.2. Word Intrusion task tanpa Stem

**Tabel 5.9 Deret Akumulatif Topik #0**

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
surabaya	0.014	14	14	0-14
indonesia	0.008	8	22	15-22
orang	0.007	7	29	23-29
suara	0.006	6	35	30-35
news	0.006	6	41	36-41
anak	0.006	6	47	42-47
foto	0.006	6	53	48-53
rumah	0.005	5	58	54-58
kawan	0.005	5	63	59-63
ciri	0.004	4	67	64-67
jatim	0.003	3	70	68-70
warga	0.003	3	73	71-73
presiden	0.003	3	76	74-76
keluarga	0.003	3	79	77-79
pk	0.003	3	82	80-82
tim	0.003	3	85	83-85
jawa	0.003	3	88	86-88
gk	0.003	3	91	89-91
reporter	0.002	2	93	92-93

kepala	0.002	2	95	94-95
--------	-------	---	----	-------

**Tabel 5.10 Pengacakan dengan *Roulette wheel* pada Topik #0**

Pengacakan ke-	Angka yang muncul	Kata
1	92	reporter
	12	surabaya
	30	suara
	52	foto
	37	news
2	70	jatim
	95	kepala
	20	indonesia
	74	presiden
	7	surabaya
3	85	tim
	21	indonesia
	14	surabaya
	69	jatim
	34	suara
4	64	ciri
	28	orang
	44	anak
	77	keluarga
	71	warga
5	50	foto
	55	rumah
	73	warga
	92	reporter
	85	tim

**Tabel 5.11 Deret Akumulatif Topik #1**

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
info	0.03	30	30	0-30
truk	0.024	24	54	31-54
kecelakaan	0.021	21	75	55-75
via	0.019	19	94	76-94
foto	0.018	18	112	95-112
lokasi	0.017	17	129	113-129
mobil	0.013	13	142	130-142
motor	0.013	13	155	143-155
jalan	0.013	13	168	156-168
petugas	0.013	13	181	169-181
update	0.011	11	192	182-192
km	0.01	10	202	193-202
korban	0.009	9	211	203-211
melaporkan	0.009	9	220	212-220
kendaraan	0.009	9	229	221-229
sepeda	0.009	9	238	230-238
tol	0.008	8	246	239-246
arah	0.008	8	254	247-254
melibatkan	0.008	8	262	255-262
sidoarjo	0.008	8	270	263-270

**Tabel 5.12 Pengacakan dengan *Roulette wheel* pada Topik #1**

Pengacakan ke-	Angka yang muncul	Kata
1	135	mobil
	41	truk

	218	melaporkan
	210	korban
	181	petugas
2	64	kecelakaan
	76	via
	21	info
	191	update
	97	foto
3	182	update
	203	korban
	80	via
	152	motor
	74	kecelakaan
4	227	kendaraan
	11	info
	131	mobil
	72	kecelakaan
	128	lokasi
5	270	sidoarjo
	29	info
	249	arah
	94	via
	243	tol

Tabel 5.13 Deret Akumulatif Topik #2

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
padat	0.07	70	70	0-70
arah	0.068	68	138	71-138
macet	0.043	43	181	139-181
jalur	0.03	30	211	182-211
imbas	0.025	25	236	212-236
tol	0.024	24	260	237-260

via	0.017	17	277	261-277
surabaya	0.017	17	294	278-294
foto	0.016	16	310	295-310
waru	0.015	15	325	311-325
simpang	0.012	12	337	326-337
jalan	0.011	11	348	338-348
mogok	0.011	11	359	349-359
raya	0.009	9	368	360-368
truk	0.009	9	377	369-377
kawan	0.009	9	386	378-386
arus	0.009	9	395	387-395
volume	0.009	9	404	396-404
sidoarjo	0.008	8	412	405-412
exit	0.008	8	420	413-420

**Tabel 5.14 Pengacakan dengan *Roulette wheel* pada Topik #2**

Pengacakan ke-	Angka yang muncul	Kata
1	340	jalan
	119	arah
	352	mogok
	244	tol
	207	jalur
2	196	jalur
	349	mogok
	8	padat
	287	surabaya
	81	arah
3	413	exit
	252	tol
	215	imbas
	66	padat
	402	volume



4	412	sidoarjo
	390	arus
	159	macet
	322	waru
	111	arah
5	404	volume
	199	jalur
	221	imbas
	46	padat
	395	arus

Tabel 5.15 Deret Akumulatif Topik #3

Distribusi Kata	Probabilitas	Probabilitas X 1000	Deret Akumulatif	<i>Roulette wheel Range</i>
arah	0.073	73	73	0-21
padat	0.069	69	142	22-36
macet	0.044	44	186	37-50
jalur	0.03	30	216	51-64
tol	0.027	27	243	65-76
imbas	0.024	24	267	77-87
via	0.021	21	288	88-95
foto	0.02	20	308	96-103
truk	0.017	17	325	104-110
waru	0.016	16	341	111-117
lintas	0.012	12	353	118-124
simpang	0.012	12	365	125-131
surabaya	0.012	12	377	132-137
jalan	0.012	12	389	138-143
lajur	0.012	12	401	144-149
mogok	0.01	10	411	150-155
waspada	0.01	10	421	156-160

arus	0.01	10	431	161-165
raya	0.009	9	440	166-170
sidoarjo	0.009	9	449	171-175

**Tabel 5.16 Pengacakan dengan *Roulette wheel* pada Topik #3**

Pengacakan ke-	Angka yang muncul	Kata
1	67	pagi
	54	wib
	38	hujan
	145	kota
	135	air
2	2	surabaya
	153	rp
	126	tiket
	116	radio
	104	kawasan
3	167	wilayah
	107	kawasan
	38	hujan
	64	wib
	98	cuaca
4	96	cuaca
	94	banjir
	157	jalan
	144	kota
	42	deras
5	140	pkl
	4	surabaya
	160	jalan
	148	kota
	152	rp

### 5.6.2.2. *Topic Intrusion task*

Sesuai dengan langkah-langkah penyusunan materi kuesioner *Topic Intrusion Task*, implementasi kuesioner dilakukan dengan mempersiapkan opsi kuesioner terlebih dahulu dengan menyusun opsi yang terdiri dari distribusi kata dalam topik, sehingga jumlah opsi akan sama dengan jumlah topik. kemudian dokumen diklasifikasikan ke dalam topik-topik sesuai probabilitas tertingginya. Dokumen-dokumen yang dipilih untuk dicantumkan ke dalam kuesioner merupakan dokumen dengan nilai probabilitas dominan lebih dari 90% untuk memastikan kecenderungannya. Kemudian setiap dokumen tersebut dipetakan dengan opsi yang ada

#### 5.6.2.2.1. *Topic Intrusion task dengan Stem*

Daftar Opsi Kuesioner *Topic Intrusion task* dengan *Stem* Eksperimen 1 dapat dilihat pada tabel Tabel 5.17

**Tabel 5.17 Daftar Opsi Kuesioner *Topic Intrusion task* dengan *Stem***

Opsi	Distribusi Kata
A	surabaya, kawan, indonesia, suara, selamat, news, anak, wib, foto, orang, radio, ibadah, ciri, hilang, kumandang, adzan, tunai, jatim, tiket, rumah
B	hujan, surabaya, pagi, kawan, jalan, jl, banjir, selamat, kawasan, air, cuaca, wib, kota, deras, daerah, gubernur, hati, listrik, kerja, suhu
C	info, celaka, lokasi, motor, truk, foto, kendaraan, via, mobil, tugas, lapor, korban, libat, jalan, bakar, update, sepeda, data, kronologi, dengar
D	arah, padat, macet, jalur, tol, imbas, via, foto, truk, waru, lintas, simpang, surabaya, jalan, lajur, mogok, waspada, arus, raya, sidoarjo

**Tabel 5.18 Daftar Opsi Kuesioner *Topic Intrusion task* dengan Stem Eksperimen 2 untuk Topik #0**

Pengacakan ke-	Angka yang muncul	Kata
1	95	kumandang
	73	orang
	64	wib
	80	radio
	102	tunai
	19	surabaya
	111	tiket
2	55	news
	89	ciri
	31	kawan
	60	anak
	76	orang
	6	surabaya
	79	radio
3	45	suara
	90	hilang
	71	foto
	77	orang
	64	wib
	93	hilang
	97	kumandang
4	96	kumandang
	110	tiket
	81	radio
	26	kawan
	103	tunai
	43	suara
	38	indonesia
5	84	ibadah
	97	kumandang
	61	anak
	73	orang
	63	wib

	98	adzan
	18	surabaya

**Tabel 5.19 Daftar Opsi Kuesioner *Topic Intrusion task* dengan *Stem* Eksperimen 2 untuk Topik #1**

Pengacakan ke-	Angka yang muncul	Kata
1	102	selamat
	159	daerah
	79	jl
	35	pagi
	11	hujan
	138	wib
	165	gubernur
2	34	pagi
	75	jalan
	131	cuaca
	145	kota
	101	selamat
	157	deras
	127	air
3	68	jalan
	95	banjir
	99	selamat
	56	kawan
	110	kawasan
	159	daerah
	157	deras
4	83	jl
	112	kawasan
	156	deras
	59	kawan
	140	wib
	9	hujan
	182	kerja

5	150	kota
	181	kerja
	131	cuaca
	77	jl
	17	hujan
	140	wib
	173	hati

**Tabel 5.20 Daftar Opsi Kuesioner *Topic Intrusion task* dengan *Stem* Eksperimen 2 untuk Topik #2**

Pengacakan ke-	Angka yang muncul	Kata
1	215	libat
	228	bakar
	134	kendara
	11	info
	102	truk
	246	sepeda
	64	lokasi
2	46	celaka
	107	truk
	70	lokasi
	170	tugas
	240	update
	235	bakar
	118	foto
3	173	tugas
	49	celaka
	79	motor
	272	kronologi
	75	lokasi
	111	foto
	251	sepeda
4	270	kronologi
	114	foto

	1	info
	65	lokasi
	201	korban
	238	update
	48	celaka
5	281	dengar
	11	info
	235	bakar
	155	mobil
	118	foto
	69	lokasi
	173	tugas

**Tabel 5.21 Daftar Opsi Kuesioner *Topic Intrusion task* dengan *Stem* Eksperimen 2 untuk Topik #3**

Pengacakan ke-	Angka yang muncul	Kata
1	309	truk
	57	arah
	186	macet
	382	jalan
	373	surabaya
	241	tol
	199	jalur
2	304	foto
	31	arah
	283	via
	102	padat
	267	imbas
	440	raya
	358	simpang
3	119	padat
	297	foto
	223	tol
	394	lajur

	95	padat
	385	jalan
	338	waru
4	272	via
	133	padat
	332	waru
	216	jalur
	314	truk
	431	arus
	295	foto
5	343	lintas
	288	via
	177	macet
	100	padat
	340	waru
	438	raya
	218	tol

**Tabel 5.22 Daftar Dokumen *Topic Intrusion task* dengan *Stem***

No	Topik	Dokumen
1	t3	07.57:1. Simpang tiga Lengkong arah Pasar Brangkal PADAT, imbas truk mogok di lajur kiri setelah pasar;2. Update Kletek, arus arah Surabaya masih PADAT imbas ada truk muat kontainer mogok di lajur kanan mendekati TL;3. Duduk Sampeyan PADAT DUA ARAH. Foto: Bara N via e100;4. Jembatan Branjangan PADAT DUA ARAH. (odp-hm)



2	t2	<p>23.55 Mobil menabrak pohon di jl Ngaglik Surabaya sebelah timur SPBU tadi sekitar pk 22.25 WIB. Mobil Isuzu Panther warna silver nopol L 1151 EF dikendarai seorang perempuan usia sekitar 65 tahun. Sebelum menabrak pohon, mobil panther itu lebih dahulu mengenai kendaraan di depannya yaitu mobil carry warna putih nopol L 8137 DI dan motor mio putih nopol L 5509 FZ. Diduga pengemudi mobil panther tsb. tidak berkonsentrasi dalam berkendara karena kecapekan. Tidak ada korban jiwa maupun luka, petugas sedang melakukan penyelidikan kecelakaan ini. Update informasi dari polsek Simokerto. Foto dikirim @MsCynthiaOei lewat @e100ss. (gkssfm)</p>
3	t2	<p>12.56: #InfoAwal: Truk menabrak rumah warga di jalan Karangnongko, Sukodono, Sidoarjo. Selain menabrak rumah, truk nopol S 8842 UP ini juga menabrak mobil nopol W 1079 YB yang sedang diparkir. Belum diketahui kronologi kejadian. Foto: Gajee via e100. (odp-wi)</p>
4	t3	<p>RT @chris_herijanto: @e100ss pemandangan macet di tol waru arah sidoarjo km 23..jpg arah sebaliknya...mrambat <a href="https://t.co/bzqHlZHU9V">https://t.co/bzqHlZHU9V</a></p>
5	t1	<p>RT @Youthssfm: Pagi yupSS.. Cuaca Surabaya pagi ini diperkirakan hujan dengan suhu 25-33 Â°C. Hope you'll have an amazing day! <a href="https://t.co/câ€">https://t.co/câ€</a></p>

6	t0	#SSInfo: Kejaksaan Tinggi Jawa Timur menyiapkan 'Sprindik' atau Surat Perintah Penyidikan pada La Nyalla Mattalitti, tersangka gratifikasi dana bank Jatim. Maruli Hutagalung, Kepala Kejati Jawa Timur mengatakan, sprindik diajukan kembali sesudah Pengadilan Negeri Surabaya mengugurkan status La Nyalla. Sebelumnya La Nyalla mangkir ke 3 kali saat Kejati Jatim melakukan pemanggilan. (news/odp-hm)
7	t0	12.32: Seorang perempuan tua a.n. Kaminem (65) yang dilaporkan hilang oleh Sumadi, anaknya, ditemukan berada oleh Liponsos Keputih. Susi, petugas Liponsos Keputih melaporkan, Kaminem ditemukan di Bangkingan oleh warga pk. 22.00 WIB semalam. Kemudian warga mengantarkan ke Mapolsek Lakarsantri yang berlanjut menyerahkan ke Liponsos Keputih. Dilaporkan, Sumadi kehilangan Ibundanya kemarin, meninggalkan rumah di daerah Perum Driyorejo. (odp-hm)
8	t3	RT @siswantoam: @e100ss sbml pintul tol kebomas terjadi antrian imbas dr truk tangki mogok di jalur kiri sbml terminal bunder gresik
9	t1	RT @dessypras: @e100ss ..daerah bendul merisi cuaca skrg mendung gelap. Mgkn sebentar lg turun hujan.
10	t1	RT @arleneflorenia: 16.33 kawasan A.Yani diguyur hujan ckp deras.. Jarak pandang terbatas..harap berhati hati @e100ss lalin pdt bergerak hâ€¦

11	t2	<p>14.23: Update #kecelakaan di Tanjungsari, pengendara sepeda motor perempuan, meninggal dan sekarang jenazah sudah dibawa ke RS PHC. AKP Bambang Sudiarto, Kasat Lantas Polres Pelabuhan Tanjung Perak melaporkan, korban a.n. Olivia (32), warga Perum KCV RI No. 72 Pakal Surabaya. Pekerjaan guru dan sekarang polisi sedang berusaha menghubungi keluarganya. Kronologinya: sepeda motor Honda Revo hitam biru W 4715 SL melaju dari barat ke timur. Saat di lokasi, motor menghindari genangan air, ngerem, oleng dan terpeleset hingga akhirnya masuk ke bawah dump truk S 8534 UP dan terlindas ban belakang. Sopir truk a.n. Itok Hermawan Widayanto, warga Sumengko Jatirejo. AKP Bambang juga menambahkan, korban bawa bungkusan besar di depan dan membatasi gerakannya. Foto: Achmad Junaedi via e100. (odp-hm)</p>
12	t3	<p>11.39: Waspada kepadatan di jalur-jalur ini, kawan. 1. Karangploso arah Pendem Macet. Foto: Kabol via e100. 2. Truk Pertamina mogok di lajur kanan Flyover Lawang. Lalu lintas padat. Foto: Sulaiman via @e100ss. (odp-pr)</p>
13	t0	<p>#SSInfo: Film animasi "Zootopia" berada di puncak box office mendapat 73,7 juta dolar. Ini hasil penjualan tiket penayangan perdana di Amerika Utara. Zootopia berhasil mengungguli kartun laris Disney 2013 lalu yakni "Frozen" yang mengumpulkan 67,4 juta dolar. Foto: movies.disney.com. (news/odp-wi)</p>

14	t3	14.30: Waspada! kepadatan di 3 jalur ini, kawan. 1. Simpang 4 Tembaan Macet. Foto: Deddy via @e100ss.2. Kletek arah Krian Macet. Foto: Beny via @e100ss.3. Balen Bojonegoro 2 arah Macet imbas kecelakaan Bus. Foto: Antok via @e100ss. (odp-pr)
15	t2	03.20 Info awal: terjadi kebakaran satu rumah di jl Manyar Tirtoasri X Surabaya. Petugas pemadam kebakaran masih melakukan upaya pemadaman. 2 unit mobil damkar dikerahkan di lokasi. Data detail belum diketahui. (gkssfm)
16	t1	RT @eeggaaaa: @e100ss Wilayah Delta sari - Pabean Asri listrik padam nih dari jam 7an td... Minggu2 mati listrik, engga asik nih @pln_123â€¦
17	t0	Anda ada disekitar Surabaya Selatan atau Surabaya Tengah? Bisa membeli tiket #JTF2016 di Radio Suara Surabaya JL Wonokitri Besar No 40 C Surabaya sampai pk 21.00 nanti. Katagori Tiket Mahasiswa/Pelajar (syarat membawa kartu identitas mahasiswa/pelajar):- Rp100 ribu per day- Rp200 ribu all dayKatagori Tiket Umum (syarat membawa kartu identitas KTP/SIM/Paspor):- Rp125 ribu per day- Rp250 ribu all day. (odp-rt)
18	t2	Foto-foto kondisi asap hitam dan api terlihat dilokasi kebakaran gudang 2 Petrokimia Gresik. Sementara, 1 unit PMK dan 1 unit mobil supply melakukan pemadaman dilokasi. Foto : Alley dan Hendro Pitono via @e100ss. (odp-rt)

19	t0	Kawan, untuk keakuratan berita mohon anda menampilkan foto2 anda sendiri, bukan broadcast/kiriman berantai. Jika ini foto orang lain tambahkan keterangan dalam gambar, Terimakasih. (odp-rs)
20	t1	#SSinfo : Banjir melanda sejumlah kecamatan di Kabupaten Aceh Selatan menyusul hujan lebat yang terjadi sepanjang Selasa (27/10/2015) dini hari. Erwiandi Kepala Pelaksana Badan Penanggulangan Bencana Daerah (BPBD) Aceh Selatan menjelaskan, leboh dari seribu rumah tergenang banjir dengan ketinggian mencapai sepinggang orang dewasa. (ssnet.rt)

#### 5.6.2.2.2. *Topic Intrusion task tanpa Stem*

**Tabel 5.23 Daftar Opsi Kuesioner *Topic Intrusion task tanpa Stem* Eksperimen 1**

Opsi	Distribusi Kata dalam Topik
A	surabaya, indonesia, orang, suara, news, anak, foto, rumah, kawan, ciri, jatim, warga, presiden, keluarga, pk, tim, jawa, gk, reporter, kepala
B	info, truk, kecelakaan, via, foto, lokasi, mobil, motor, jalan, petugas, update, km, korban, melaporkan, kendaraan, sepeda, tol, arah, melibatkan, sidoarjo
C	padat, arah, macet, jalur, imbas, tol, via, surabaya, foto, waru, simpang, jalan, mogok, raya, truk, kawan, arus, volume, sidoarjo, exit
D	surabaya, kawan, hujan, wib, pagi, selamat, banjir, cuaca, kawasan, radio, jl, tiket, air, pkl, kota, rp, jalan, deras, wilayah, pk

**Tabel 5.24 Daftar Opsi Kuesioner *Topic Intrusion task* tanpa *Stem*  
Eksperimen 2 Topik #0**

Pengacakan ke-	Angka yang muncul	Kata
1	16	indonesia
	46	anak
	73	warga
	8	surabaya
	34	suara
	26	orang
	50	foto
2	67	ciri
	33	suara
	63	kawan
	37	news
	72	warga
	57	rumah
	13	surabaya
3	73	warga
	19	indonesia
	35	suara
	61	kawan
	21	indonesia
	23	orang
	77	keluarga
4	93	reporter
	22	indonesia
	53	foto
	81	pk
	24	orang
	62	kawan
	88	jawa
5	65	ciri
	41	news
	27	orang
	93	reporter
	57	rumah

	92	reporter
	74	presiden

**Tabel 5.25 Daftar Opsi Kuesioner *Topic Intrusion task* tanpa *Stem*  
Eksperimen 2 Topik #1**

Pengacakan ke-	Angka yang muncul	Kata
1	216	melaporkan
	109	foto
	66	kecelakaan
	230	sepeda
	145	motor
	170	petugas
	46	truk
2	243	tol
	55	kecelakaan
	260	melibatkan
	163	jalan
	123	lokasi
	248	arah
	130	mobil
3	17	info
	211	korban
	168	jalan
	52	truk
	195	km
	260	melibatkan
	118	lokasi
4	122	lokasi
	74	kecelakaan
	152	motor
	219	melaporkan
	127	lokasi
	5	info
	151	motor

5	26	info
	149	motor
	82	via
	39	truk
	210	korban
	216	melaporkan
	256	melibatkan

**Tabel 5.26 Daftar Opsi Kuesioner *Topic Intrusion task* tanpa *Stem*  
Eksperimen 2 Topik #2**

Pengacakan ke-	Angka yang muncul	Kata
1	141	macet
	250	tol
	189	jalur
	331	simpang
	20	padat
	343	jalan
	123	arah
2	191	jalur
	396	volume
	272	via
	108	arah
	288	surabaya
	351	mogok
	57	padat
3	161	macet
	398	volume
	337	simpang
	371	truk
	106	arah
	405	sidoarjo
	35	padat
4	239	tol
	119	arah



	268	via
	151	macet
	209	jalur
	386	kawan
	17	padat
5	298	foto
	12	padat
	260	tol
	177	macet
	278	surabaya
	356	mogok
	305	foto

**Tabel 5.27 Daftar Opsi Kuesioner *Topic Intrusion task* tanpa Stem Eksperimen 2 Topik #3**

Pengacakan ke-	Angka yang muncul	Kata
1	30	kawan
	2	surabaya
	173	pk
	143	pkl
	58	wib
	152	rp
	43	hujan
2	149	kota
	132	air
	126	tiket
	53	wib
	158	jalan
	88	banjir
	17	surabaya
3	49	hujan
	78	selamat
	130	tiket
	138	pkl

	7	surabaya
	92	banjir
	172	pk
4	150	rp
	132	air
	40	hujan
	67	pagi
	131	tiket
	69	pagi
	59	wib
5	131	tiket
	108	kawasan
	77	selamat
	17	surabaya
	170	wilayah
	32	kawan
	121	jl

**Tabel 5.28 Daftar Dokumen *Topic Intrusion task* tanpa *Stem***

No	Kunci Jawaban	Dokumen
----	---------------	---------

1	t0	<p>#Sstoday : "Jurus Ekonomi Jokowi"Kemarin, Jokowi Presiden mengeluarkan kebijakan ekonomi yang satu diantaranya berfokus meningkatkan daya beli masyarakat yang turun akibat penguatan dollar. Bambang Brodjonegoro Menteri Keuangan mengatakan, kebijakan itu antara lain menaikkan Penghasilan Tidak Kena Pajak (PTKP) pada pertengahan tahun menjadi Rp36,3 juta per tahun, tax Holiday untuk industri pioner, tax Allowance untuk mendorong sektor industri pengolahan. Untuk UMKM, pemerintah menugaskan secara khusus Lembaga Pembiayaan Ekspor Indonesia (LPEI) dan mencegah UMKM melakukan pemutusan hubungan kerja (PHK) terhadap karyawannya. Dalam waktu dekat, pemerintah juga akan menyelesaikan Peraturan Pemerintah (PP) mengenai Pajak Pertambahan Nilai (PPn) tidak dipungut biaya untuk sektor galangan kapal.Simak selengkapnya di FM 100 Suara Surabaya. (odp-rt)</p>
2	t1	<p>13.53: Update #kebakaran di Desa Candi Negero Wonoayu Sidoarjo. Bangunan terbakar 1 unit rumah, milik Bapak Agus. Penyebab kebakaran karena konsleting listrik. Ada 2 unit PMK di lokasi. Bripka Joko - Anggota Polsek Wonoayu mengatakan, warga panik karena rumah disini saling berhimpitan. Angin yang cukup kencang membuat api cukup besar, tetapi tidak sampai merembet ke bangunan lain. Kondisi rumah hangus 50%. Saat ini api sudah berhasil dipadamkan. (odp-pr)</p>

3	t3	<p>#JTF2015 Mau dapat tiket Jazz Traffic Festival 2015 harga presale 1? Datang ke Artisan Market Graha Fair Ground, Jl. Mayjend Yono Suwoyo, Surabaya Barat sekarang. Harga tiket #JTF2015 presale 1 katagori UMUM Rp 120.000/day dan Rp 240.000 all day. Katagori Mahasiswa, tiket #JTF2015 masih diharga Rp 100.000/day dan Rp 200.000 all day. Jangan lupa bawa KTM ya. Ini adalah kesempatan terakhir kalian untuk mendapatkan tiket #JTF2015 dengan harga PRESALE 1 ya! Jadi, manfaatkan kesempatan ini. (odp-pr)</p>
4	t0	<p>#SSToday Ini tahun ke-3 dari program Pelajar Pelopor. DR Ikhsan. S.Psi. MM - Kepala Dinas Pendidikan Surabaya mengatakan, meski masih SMP atau SMA, kegiatan mereka sangat positif. Seperti membuat perpustakaan di kampungnya untuk membantu anak-anak agar hobi membaca. Ada juga yang belajar jadi pengusaha udang, menularkan ilmu bisnis kepada teman-temannya. Hal ini merupakan potensi yang bagus, harus dijadikan contoh dan motivasi bagi yang lain. Jika hal seperti ini tidak beri wadah, potensi-potensi yang luar biasa seperti ini akan mati. (odp-pr)</p>
5	t2	<p>07.34: Waspada kepadatan di 3 jalur ini,kawan. 1. Sukodono - Kletek arah Surabaya padat imbas kegiatan jalan sehat.2. Sarirogo padat imbas kegiatan jalan sehat.3. Brigjen Katamso arah Rewwin padat imbas kegiatan jalan sehat.(odp-pr)</p>

6	t2	<p>17.41: Jalur-jalur ini padat, imbas volume tinggi dan kendaraan bermasalah.1. Flyover Trosobo arah Mojokerto padat imbas Dump Truk Mogok di lajur kiri pas tanjakan.2. Jelang Exit Tol Sidoarjo padat, ada 2 Truk bermasalah di bahu jalan.3. Truk bermasalah di depan RS Citra Medika lajur kiri arah Jombang. Ekor antrean hingga SPBU Tjiwi. Foto: Samuel via e100.4. Sebelum TL Jetak arah Surabaya padat. Foto: Santoso via @e100ss. 5. Gunungsari arah Kedurus padat. (odp-pr)</p>
7	t1	<p>17.27: Kecelakaan melibatkan motor gede (Nopol tidak terapanantau oleh Polisi karena saat Polisi datang moge sudah dibawa oleh rombongan lainnya dengan mobil) dan sepeda motor Honda Vario AG 5153 XC. Bripka Edy Anggota Lantas Polres Nganjuk mengatakan, lokasi tepatnya di Raya Baron - Kertosono sebelah barat Wisata Waterpark dan kecelakaan terjadi pk 14.00. Kronologi kejadian, moge dan sepeda motor vario ini berjalan searah, Nganjuk - Surabaya. Moge menyerempet vario, dua penumpang vario sepasang suami istri jatuh sedangkan moge terus melaju sampai menabrak pohon dan ban bagian depannya masuk selokan. Pengendara moge, pengendara vario dan yang dibonceng mengalami luka ringan dibawa ke RS Sumber Waras. Saat ini dilokasi sudah clear. (odp-rt)</p>

8	t3	07.13: Selamat pagi kawan. Beberapa kawasan Surabaya pagi ini yang dilaporkan Hujan. Terpantau hujan mengguyur Benowo, Tandes, dan hujan ringan di Mayjen Sungkono. Selamat menikmati akhir pekan kawan. "Bakat terbentuk dalam gelombang kesunyian, watak terbentuk dalam riak besar kehidupan" - Goethe (odp-pr)
9	t1	13.03: Info awal dua #kecelakaan:1. Jelang Layang Trosobo arah Krian, melibatkan sepeda motor dengan Elf. Didin, pendengar SS melaporkan, satu orang terluka dan tergeletak di pinggir jalan. Sudah ditolong warga. Belum diketahui data dan kronologinya. Info sudah diteruskan ke petugas. Foto: Denny D via e100;2. Perum Bukit Darmo Golf, melibatkan mobil Agya nopol L 1062 BK dengan Pajero hitam. Tidak ada korban jiwa, sementara bagian depan mobil ringsek. Foto: Theresia S via @e100ss. (odp-hmn)
10	t2	20.43:1. Update Pandaan, arus arah Surabaya masih PADAT imbas dump truk mogok di lajur kanan TL patung sapi;2. Raya Klakah Lumajang, arus arah Surabaya MACET imbas pengaspalan. Setelah itu ada truk mogok di lajur kiri Raya Ranuyoso. Sudah ada pengaturan dari petugas. (odp-hm)

11	t0	<p>23.40 Ditemukan orang tua: mengaku bernama ibu Wagini atau dipanggil Cikrak. Usia sekitar 90 tahun. Ciri-ciri kulit sawo matang, pakai kacamata, pakai baju lengan panjang warna hijau muda, jilbab biru, rok warna hijau. Beliau membawa tas hitam dan tas plastik warna merah. Ibu Wagini ditemukan di warung pojok sekitar bundaran Dolog. Ketika ditanyai beberapa orang di warung tsb., ibu Wagini mengaku ingin pergi ke rumah anaknya di daerah Jemursari tapi sayangnya lupa alamat lengkap anaknya. Ibu Wagini tidak membawa kartu identitas diri. Saat ini posisi ibu Wagini dititipkan di mapolsek Wonocolo. (gkssfm)</p>
12	t3	<p>#InfoSAMSAT: Jadwal Samsat Keliling Wilayah PoLDA Jatim Senin 3 Agustus 20151. Polrestabes Surabaya :a. Samsat Corner Royal Plaza pkl 10.00 s/d 20.00 Wib.b. Samsat Corner Carrefour Rungkut pkl 10.00 s/d 20.00 Wib.c. Samsat Corner ITC pkl 10.00 s/d 20.00 Wib.d. Samsat Corner PTC, pkl 10.00 s/d 20.00 Wib.e. Samsat Corner Galaxy Mall pkl 10.00 s/d selesai.f. Samsat Corner dan ATM Samsat di Grand City Pkl 10.00 s/d 20.00 Wib. Pelayanan Samsat Keliling hanya untuk pembayaran pajak kendaraan bermotor 1 tahunan dan pengesahan STNK. (odp-hm)</p>

13	t2	18.32:1. Balongbendo arah Mojokerto MACET, imbas pengerjaan jalan dengan alat berat. Ekor antrean sejak depan SPBU. Arus didominasi kendaraan besar;2. Update Benowo, arus dari Kepatihan DITUTUP dan dialihkan ke Gelora Bung Tomo, imbas truk yang mogok sejak TIGA JAM lalu;3. HR Muhammad PADAT, simpulnya di putar balik patung kuda. (odp-hm)
14	t3	#SSInfo Sejak kemarin siang sampai semalam memang Sidoarjo diguyur hujan deras. Fathurahman - Kepala Dinas PU Pengairan Sidoarjo mengatakan, banjir sudah berkurang dari kemarin yang paling dalam 60cm sekarang 20cm. Masalah utama banjir di Sidoarjo adalah Drainase. Saat ini pompa sudah digunakan, semua saluran sudah saling terkoneksi sehingga genangan bisa segera surut. (odp-pr)
15	t1	9.37 : Update : Truk muat pupuk AE 8141 UB yang terguling di Dusun Ngrajek Sambirejo Tanjung Anom Nganjuk, adalah korban kecelakaan. Aiptu Sumardi Anggota Laka Lantas Polres Nganjuk menjelaskan, truk ini mengalami kecelakaan dengan truk box kosong W 9503 M. Kronologinya, truk box dari arah Nganjuk-Surabaya dan truk gandeng dari arah sebaliknya. Diduga sopir truk box ngantuk, oleng ke kanan dan kena truk gandeng. Akibatnya kepala dan gandengan terguling. Muatan pupuk yang tumpah sudah diangkut kuli. Lalu lintas sudah normal. Foto : Wheny Purnomo via e100. Foto 2 : Oviq via e100. (odp-rt)



16	t0	<p>Mohon bantuannya, Kawan.. Morin pendengar SS kehilangan Naam Suhadi (65) kakaknya sejak pk 07.00 pagi tadi. Naam biasanya bersepeda setiap pagi dari daerah Manukan ke Kupang Krajan mengunjungi kakaknya. Tapi, hari ini saat dicek kerumah kakaknya, Naam tidak ada disana. keluarga mencoba mencari didaerah Kupang Krajan juga belum ketemu. Ciri-Ciri Naam : menggunakan training warna biru, sepatu kets, berkulit hitam, tinggi badan 170 cm, rambut putih, hidung mancung. Saat meninggalkan rumah, tidak bawa identitas apapun. Apabila ada kawan-kawan yang melihat keberadaan Naam, sila hubungi GKSSFM di (031) 5600000. (odp-rt)</p>
17	t3	<p>#InfoSIM Jadwal SIM Keliling Wilayah PoLDA Jatim, Kamis 3 September 2015;1. Polrestabes Surabaya:a. Taman Bungkul pk1 09.00 s/d 16.00 Wib.b. Pasar Tambakrejo pk1 09.00 s/d 16.00 Wibc. SIM Corner Tunjungan Plaza pk1 10.00 s/d selesai.d. SIM Corner PTC pk1 10.00 s/d selesai.2. Polres Sidoarjo di SIM Corner Sun City pk1 10.00 s/d selesai.3. Polres Kediri kota di Jl.Brawijaya pk1 08.00 s/d 12.00 wib;4. Polres Pasuruan di pasar baru Beji pk1 08.30 s/d selesai;5. Polres Bojonegoro di Bravo Swalayan pk1 08.00 s/d 12.00 Wib.SIM Keliling melayani perpanjangan SIM A dan C yang masa berlakunya habis sebelum 3 bulan. (odp-pr)</p>

18	t0	#SSinfo : Jokowi Presiden hari ini dijadwalkan meluncurkan paket kebijakan jilid 5. Pemerintah terus menerbitkan paket-paket, jangka pendek maupun menengah. Para Menteri bidang Ekonomi diminta membuat paket kebijakan, tiap 2 minggu sekali. Indonesia harus mampu bersaing dengan negara lain, seperti Vietnam dan India yang pertumbuhan ekonominya cepat. (ssnet.rt)
19	t1	09.20 : WASPADA TOL ! Info awal : Kecelakaan di Tol Waru-Satelit KM 12 melibatkan mobil pick up vs Grand Livina. Info dari Jhon pendengar SS, posisi kap grand livina masuk kebawah bak pick up. Lajur kanan tertutup total. Tim GKSSFM sudah menghubungi Senkom Jasa Marga untuk koordinasi. Foto : Egy Henry via @e100ss. (odp-rt)
20	t2	13.30 : Rangkuman jalur MACET siang ini : 1.Trailler muat kontainer mogok dilajur kiri pintu keluar Tol Waru. Foto : Jimmy Mathovani via e100;2.Mobil mogok di Tol Dupak-Waru KM 12 lajur kanan. Lalu lintas padat. Foto : Jimmy Mathovani via e100;3.Imbas kemacetan di JL Raya Panjang Jiwo, ekor kemacetannya sudah sampai Jembatan MERR Stikom - Panjang Jiwo;4.Medaeng - Krian;5.Mastrip - Karang Pilang. Ekornya di SPBU Kedurus. (odp-rt)

### 5.6.3. Sistematika Kuesioner Uji Koherensi Topik

Tujuan dalam melakukan perancangan sistematika kuesioner adalah untuk memudahkan responden dalam memahami konteks kuesioner, sehingga dapat memberikan hasil yang lebih objektif.

Sesuai dengan rancangan sistematika kuesioner uji koherensi topik, terdapat tiga bagian yaitu bagian judul, bagian deskripsi pembuka dan bagian pertanyaan. Pada bagian judul berfungsi untuk mengidentifikasi aktivitas dan metode uji koherensi topik, pada bagian deskripsi pembuka berisi pengenalan penulis, tujuan pengumpulan data, deskripsi singkat metode pengumpulan data dan pengantar cara pengisian data. Berikut merupakan implementasi kuesioner dengan *tools Google Form* yang dapat dilihat pada Gambar 2.1.

**Kuesioner Uji Koherensi Topik**

Salam Sejahtera,

Perkenalkan, saya I Made Kusnanta Bramantya Putra, mahasiswa smt 8 Departemen Sistem Informasi, Fakultas Teknologi Informasi ITS.

Dalam rangka pengerjaan Tugas Akhir dengan judul Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA), saya mohon kesediaan Bapak/Ibu/Saudara/i untuk mengisi kuesioner di bawah ini sebagai kebutuhan data.

Atas kesediaan dan perhatian yang diberikan, saya ucapkan terimakasih

Hormat saya,  
I Made Kusnanta Bramantya Putra

**BERIKUTNYA**

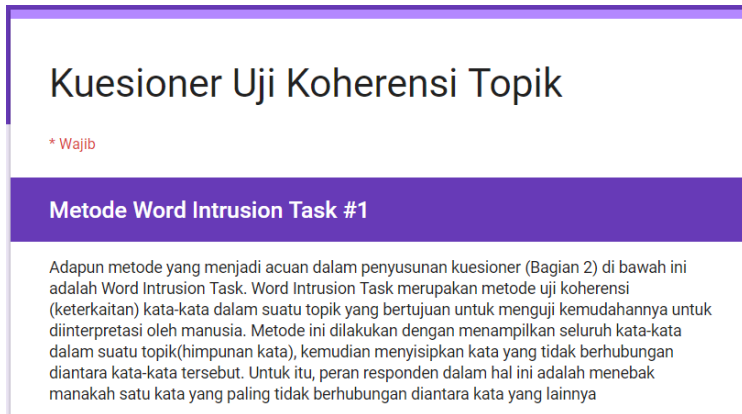
Halaman 1 dari 5

**Gambar 5.1** Bagian judul kuesioner uji koherensi topik

#### 5.6.3.1. *Word Intrusion Task*

Pada kuesioner bagian *Word Intrusion Task*, pertanyaan ditujukan agar responden memilih satu kata yang paling tidak berhubungan dari kumpulan opsi yang ditampilkan. Untuk setiap kali menampilkan kuesioner, opsi pilihan akan diacak urutannya. Berikut merupakan implementasi kuesioner

pada bagian *Word Intrusion task* yang dapat dilihat pada Gambar 5.2 dan Gambar 5.3



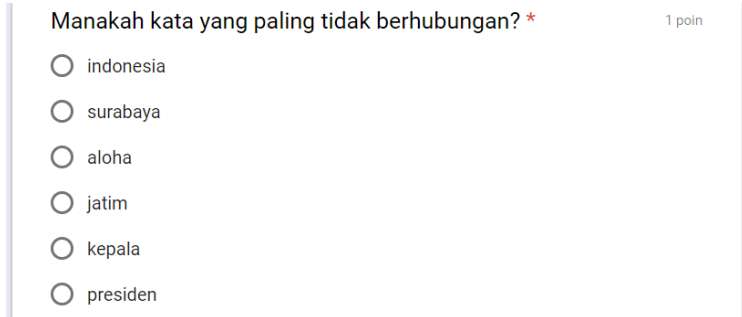
**Kuesioner Uji Koherensi Topik**

\* Wajib

**Metode Word Intrusion Task #1**

Adapun metode yang menjadi acuan dalam penyusunan kuesioner (Bagian 2) di bawah ini adalah Word Intrusion Task. Word Intrusion Task merupakan metode uji koherensi (keterkaitan) kata-kata dalam suatu topik yang bertujuan untuk menguji kemudahannya untuk diinterpretasi oleh manusia. Metode ini dilakukan dengan menampilkan seluruh kata-kata dalam suatu topik(himpunan kata), kemudian menyisipkan kata yang tidak berhubungan diantara kata-kata tersebut. Untuk itu, peran responden dalam hal ini adalah menebak manakah satu kata yang paling tidak berhubungan diantara kata yang lainnya

**Gambar 5.2** Bagian deskripsi pembuka kuesioner uji koherensi topik bagian *Word Intrusion Task*



Manakah kata yang paling tidak berhubungan? \* 1 poin

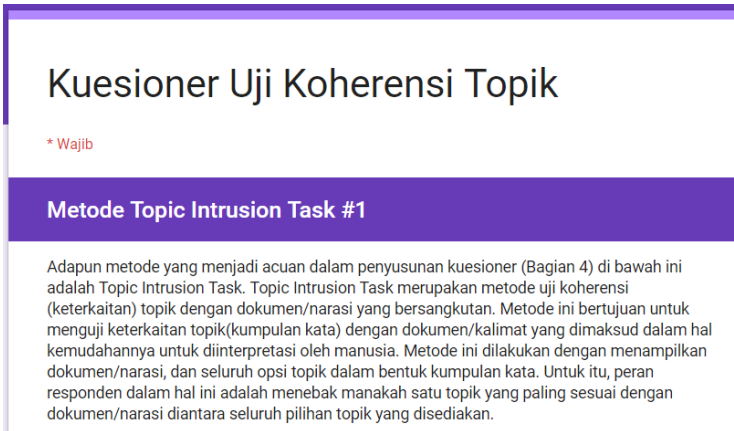
- ☐ indonesia
- ☐ surabaya
- ☐ aloha
- ☐ jatim
- ☐ kepala
- ☐ presiden

**Gambar 5.3** Bagian pertanyaan kuesioner uji koherensi topik bagian *Word Intrusion Task*

### 5.6.3.2. *Topic Intrusion Task*

Pada kuesioner bagian *Topic Intrusion Task*, pertanyaan ditujukan agar responden memilih satu topik (himpunan kata) yang paling berhubungan dengan satu dokumen yang ditampilkan. Untuk setiap kali menampilkan

kuesioner, opsi pilihan tidak diacak urutannya. Berikut merupakan implementasi kuesioner pada bagian *Topic Intrusion task* yang dapat dilihat pada Gambar 5.4 sampai Gambar 5.6



**Kuesioner Uji Koherensi Topik**

\* Wajib

**Metode Topic Intrusion Task #1**

Adapun metode yang menjadi acuan dalam penyusunan kuesioner (Bagian 4) di bawah ini adalah Topic Intrusion Task. Topic Intrusion Task merupakan metode uji koherensi (keterkaitan) topik dengan dokumen/narasi yang bersangkutan. Metode ini bertujuan untuk menguji keterkaitan topik(kumpulan kata) dengan dokumen/kalimat yang dimaksud dalam hal kemudahannya untuk diinterpretasi oleh manusia. Metode ini dilakukan dengan menampilkan dokumen/narasi, dan seluruh opsi topik dalam bentuk kumpulan kata. Untuk itu, peran responden dalam hal ini adalah menebak manakah satu topik yang paling sesuai dengan dokumen/narasi diantara seluruh pilihan topik yang disediakan.

**Gambar 5.4** Bagian deskripsi pembuka kuesioner uji koherensi topik bagian *Topic Intrusion Task*

Manakah topik(himpunan kata) yang paling sesuai dengan kumpulan teks berikut? \*

1 poin

13.53: Update #kebakaran di Desa Candi Ngoro Wonoayu Sidoarjo. Bangunan terbakar 1 unit rumah, milik Bapak Agus. Penyebab kebakaran karena konsleting listrik. Ada 2 unit PMK di lokasi. Bripka Joko - Anggota Polsek Wonoayu mengatakan, warga panik karena rumah disini saling berhimpitan. Angin yang cukup kencang membuat api cukup besar, tetapi tidak sampai merembet ke bangunan lain. Kondisi rumah hangus 50%. Saat ini api sudah berhasil dipadamkan. (odp-pr)

- ☐ topic #0 - surabaya , indonesia , orang , suara , news , anak , foto , rumah , kawan , ciri , jatim , warga , presiden , keluarga , pk , tim , jawa , gk , reporter , kepala
- ☐ topic #1 - info , truk , kecelakaan , via , foto , lokasi , mobil , motor , jalan , petugas , update , km , korban , melaporkan , kendaraan , sepeda , tol , arah , melibatkan , sidoarjo
- ☐ padat , arah , macet , jalur , imbas , tol , via , surabaya , foto , waru , simpang , jalan , mogok , raya , truk , kawan , arus , volume , sidoarjo , exit
- ☐ surabaya , kawan , hujan , wib , pagi , selamat , banjir , cuaca , kawasan , radio , jl , tiket , air , pk , kota , rp , jalan , deras , wilayah , pk

**Gambar 5.5** Bagian pertanyaan kuesioner uji koherensi topik bagian *Topic Intrusion task* Eksperimen 1

Manakah topik(himpunan kata) yang paling sesuai dengan kumpulan teks berikut? \*

1 poin

07.34: Waspadai kepadatan di 3 jalur ini,kawan. 1. Sukodono - Kletek arah Surabaya padat imbas kegiatan jalan sehat. 2. Sarirogo padat imbas kegiatan jalan sehat. 3. Brigjen Katamsa arah Rewwin padat imbas kegiatan jalan sehat.(odp-pr)

- ☐ info , motor , via , truk , korban , melaporkan , melibatkan
- ☐ foto , padat , tol , macet , surabaya , mogok , foto
- ☐ ciri , news , orang , reporter , rumah , reporter , presiden
- ☐ tiket , kawasan , selamat , surabaya , wilayah , kawan , jl

**Gambar 5.6** Bagian pertanyaan kuesioner uji koherensi topik bagian *Topic Intrusion task* Eksperimen 2

## BAB VI

### HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan proses pengujian dan analisis terhadap hasil pengujian yang diperoleh dari proses implementasi yang telah dibahas pada bab sebelumnya.

#### 6.1. Mempersiapkan Data

##### 6.1.1. *Loading Data*

Jumlah data mentah yang dimuat sebagai sumber data masukan dapat dilihat pada Tabel 6.1 Data mentah yang dimuat sebagai sumber data masukan

**Tabel 6.1 Data mentah yang dimuat sebagai sumber data masukan**

Jenis data	Jumlah data	Jumlah kata
<i>Tweet</i> masyarakat yang di <i>retweet</i> oleh e100ss	10238	157271
<i>Post Facebook</i>	27036	908744
Total data	37274	1066015

##### 6.1.2. Pembersihan Data

Jumlah data yang telah dibersihkan dari *url*, *mention*, *reserved words*, angka, karakter *non-alphanumeric* dan token sepanjang 1 digit dapat dilihat pada Tabel 6.2 Hasil pembersihan data

**Tabel 6.2 Hasil pembersihan data**

	Jumlah dokumen	Jumlah kata
Total data	37274	1033455

#### 6.2. Pra-Proses Data

##### 6.2.1. Pendefinisian *Stopwords*

*Stopwords* merupakan kata umum (*common words*) yang biasanya muncul dalam jumlah besar atau berfrekuensi

tinggi dan dianggap tidak memiliki makna. Dalam kasus ini, pendefinisian *stopword* dilakukan dengan tiga justifikasi, yaitu

1. Berdasarkan pemaknaan kata sesuai sistem tata Bahasa Indonesia baku
2. Berdasarkan analisis kemunculan kata dalam daftar intensitas kemunculan kata (1000 kata tertinggi)
3. Kode komunikasi internal e100ss

Dalam sistem tata Bahasa Indonesia baku, terdapat istilah kelas kata, yaitu istilah linguistik kelas atau golongan (kategori) kata berdasarkan bentuk, fungsi, atau maknanya (KBBI). Diantara kelas-kelas kata yang ada, terdapat beberapa kelas kata yang memiliki arti gramatikal namun tidak memiliki arti leksikal. Dalam artian lain, bahwa suatu kata jika berdiri sendiri tanpa disertai dengan kata yang diterangkan, kata tersebut tidak memiliki makna, sehingga kelas kata yang dianggap memenuhi syarat untuk digolongkan ke dalam *stopword* dapat dilihat pada Tabel 6.3 Kelas kata yang memenuhi syarat sebagai *stopword*

**Tabel 6.3 Kelas kata yang memenuhi syarat sebagai *stopword***

Kelas Kata		Definisi
Adverbia		Kata Keterangan
Pronomina		Kata Ganti
Numeralia		Kata Bilangan
Kata Tugas	Preposisi	Kata Depan
	Konjungsi	Kata Sambung
	Artikula	Kata Sandang
	Interjeksi	Kata Seru
	Partikel	Partikel

Dalam melakukan pertimbangan berdasarkan analisis kemunculan kata dalam daftar intensitas kemunculan kata (1000 kata tertinggi), dapat memperhitungkan kelas kata di luar kelas kata yang dicantumkan. Adapun kata-kata yang dihilangkan merupakan kata yang memiliki intensitas



kemunculan tinggi namun tidak memiliki makna. Hasil pendefinisian *stopword* yang digunakan dalam tahap pra-proses data dapat dilihat pada Lampiran A.1

Jumlah data yang telah dibersihkan oleh *stopwords* dapat dilihat pada Tabel 6.4 Hasil pembersihan data melalui *stopwords*

**Tabel 6.4 Hasil pembersihan data melalui *stopwords***

	Jumlah Dokumen	Jumlah kata
Total data	21364	711807

### 6.3. Pembentukan Model *LDA*

Dalam melakukan pembentukan model *LDA*, terdapat model yang dibentuk, model yang melalui tahap *Stemming* dan model yang tidak melalui tahap *Stemming*

#### 6.3.1. Hasil Pembentukan Model *LDA* dengan *Stemming*

Distribusi probabilitas kata dalam 4 Topik (20 kata per topik) dari model yang melalui proses *Stemming* dapat dilihat pada Tabel 6.5 Hasil Pembentukan Model *LDA* dengan *Stemming*

**Tabel 6.5 Hasil Pembentukan Model *LDA* dengan *Stemming***

<b>topik #0:</b>	<b>topik #1:</b>	<b>topik #2:</b>	<b>topik #3:</b>
0.021*surabaya	0.017*hujan	0.035*info	0.073*arah
0.010*kawan	0.016*surabaya	0.023*celaka	0.069*padat
0.008*indonesia	0.015*pagi	0.019*lokasi	0.044*macet
0.007*suara	0.014*kawan	0.015*motor	0.030*jalur
0.006*selamat	0.014*jalan	0.015*truk	0.027*tol
0.005*news	0.011*jl	0.015*foto	0.024*imbis
0.005*anak	0.011*banjir	0.015*kendara	0.021*via
0.005*wib	0.010*selamat	0.014*via	0.020*foto
0.005*foto	0.010*kawasan	0.014*mobil	0.017*truk
0.005*orang	0.009*air	0.013*tugas	0.016*waru
0.004*radio	0.009*cuaca	0.013*lapor	0.012*lintas
0.004*ibadah	0.008*wib	0.012*korban	0.012*simpang
0.004*ciri	0.007*kota	0.012*libat	0.012*surabaya
0.004*hilang	0.006*deras	0.010*jalan	0.012*jalan
0.004*kumandang	0.006*daerah	0.010*bakar	0.012*lajur
0.004*adzan	0.006*gubernur	0.010*update	0.010*mogok
0.004*tunai	0.005*hati	0.010*sepeda	0.010*waspada
0.004*jatim	0.005*listrik	0.009*data	0.010*arus

0.003*tiket	0.005*kerja	0.009*kronologi	0.009*raya
0.003*rumah	0.005*suhu	0.008*dengar	0.009*sidoarjo

### 6.3.2. Hasil Pembentukan Model LDA tanpa Stemming

Distribusi probabilitas kata dalam 4 Topik (20 kata per topik) dari model yang tidak melalui proses *Stemming* dapat dilihat pada Tabel 6.6 Hasil Pembentukan Model LDA tanpa *Stemming*

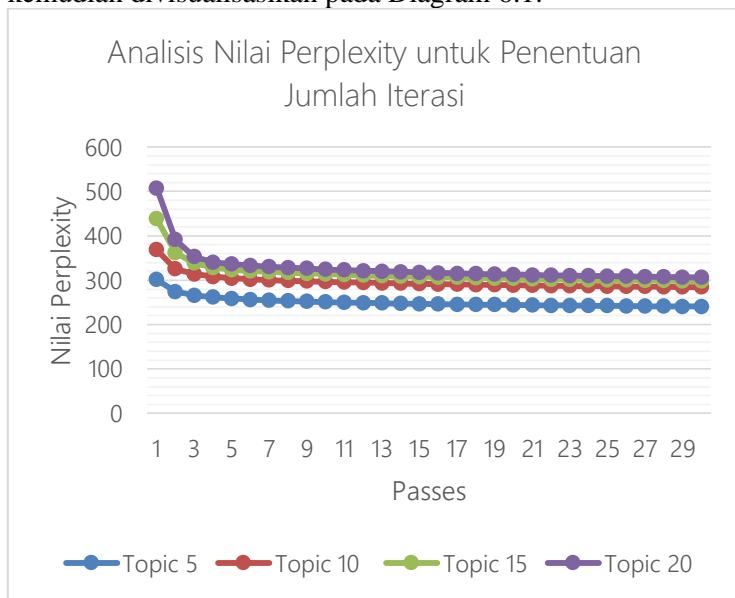
**Tabel 6.6 Hasil Pembentukan Model LDA tanpa Stemming**

topik #0:	topik #1:	topik #2:	topik #3:
0.014*surabaya	0.030*info	0.070*padat	0.021*surabaya
0.008*indonesia	0.024*truk	0.068*arah	0.015*kawan
0.007*orang	0.021*kecelakaan	0.043*macet	0.014*hujan
0.006*suara	0.019*via	0.030*jalur	0.014*wib
0.006*news	0.018*foto	0.025*imbas	0.012*pagi
0.006*anak	0.017*lokasi	0.024*tol	0.011*selamat
0.006*foto	0.013*mobil	0.017*via	0.008*banjir
0.005*rumah	0.013*motor	0.017*surabaya	0.008*cuaca
0.005*kawan	0.013*jalan	0.016*foto	0.007*kawasan
0.004*ciri	0.013*petugas	0.015*waru	0.007*radio
0.003*jatim	0.011*update	0.012*simpang	0.007*jl
0.003*warga	0.010*km	0.011*jalan	0.007*tiket
0.003*presiden	0.009*korban	0.011*mogok	0.006*air
0.003*keluarga	0.009*melaporkan	0.009*raya	0.006*pk
0.003*pk	0.009*kendaraan	0.009*truk	0.006*kota
0.003*tim	0.009*sepeda	0.009*kawan	0.006*rp
0.003*jawa	0.008*tol	0.009*arus	0.005*jalan
0.003*gk	0.008*arah	0.009*volume	0.005*deras
0.002*reporter	0.008*melibatkan	0.008*sidoarjo	0.005*wilayah
0.002*kepala	0.008*sidoarjo	0.008*exit	0.005*pk

## 6.4. Validasi Model Topik

### 6.4.1. Penentuan Jumlah Iterasi

Dalam melakukan penentuan jumlah iterasi (*passes*), metode yang digunakan adalah dengan melakukan analisis nilai *perplexity*. Analisis nilai *perplexity* dilakukan dengan cara menjalankan pemodelan terhadap topik dengan setidaknya tiga parameter topik yang berbeda dengan nilai *passes* mula-mula 30. Dalam kasus ini, parameter jumlah topik yang dipilih adalah 5, 10, 15 dan 20. Hasil nilai *perplexity* yang muncul dari masing-masing parameter jumlah topik ini kemudian dicatat kemudian divisualisasikan pada Diagram 6.1.



**Diagram 6.1 Analisis Nilai *Perplexity* untuk Penentuan Jumlah Iterasi**

Berdasarkan visualisasi Diagram 6.1, dapat dilihat bahwa nilai *perplexity* sudah mencapai kondisi cenderung stabil pada *passes* ke 10 untuk keseluruhan parameter jumlah topik, sehingga dapat disimpulkan bahwa iterasi yang digunakan adalah 10.

### 6.4.2. Penentuan Jumlah Topik

Penentuan jumlah topik dilakukan dengan menganalisis nilai *perplexity*, namun analisis nilai *perplexity* dalam konteks untuk penentuan jumlah topik dilakukan dengan melakukan eksperimen pada parameter jumlah topik dalam rentang nilai yang lebih luas, dalam hal ini rentang nilai yang digunakan dalam eksperimen ditampilkan dalam tabel di bawah. Analisis nilai *perplexity* dalam konteks untuk penentuan jumlah topik dilakukan dengan melakukan running sebanyak 30 kali untuk mendapatkan rata-rata nilai *perplexity* yang akurat untuk masing-masing parameter jumlah topik.

#### 6.4.2.1. Penentuan Jumlah Topik dengan *Stemming*

Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik dengan *Stemming* dapat dilihat pada lampiran A2

##### 6.4.2.1.1. Analisis Hasil Eksperimen Nilai *Perplexity* untuk Penentuan Jumlah Topik

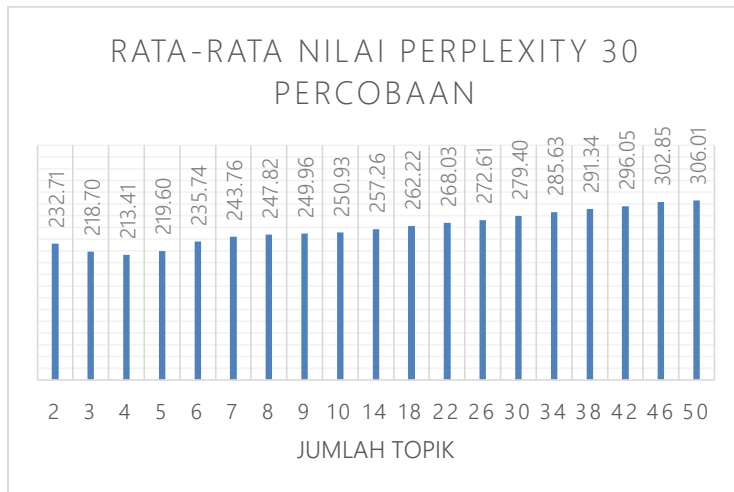
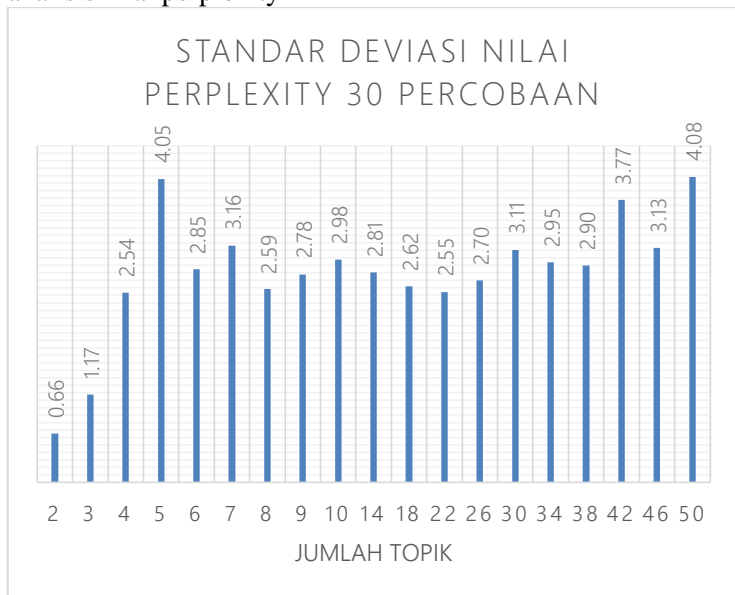


Diagram 6.2 Rata-rata nilai *Perplexity* 30 percobaan

Berdasarkan Diagram 6.2, nilai *perplexity* terendah terdapat pada jumlah topik 4 yaitu 213.41, dan tren nilai *perplexity* meningkat untuk jumlah topik yang semakin tinggi, sehingga 4 topik merupakan jumlah topik terbaik berdasarkan analisis nilai *perplexity*



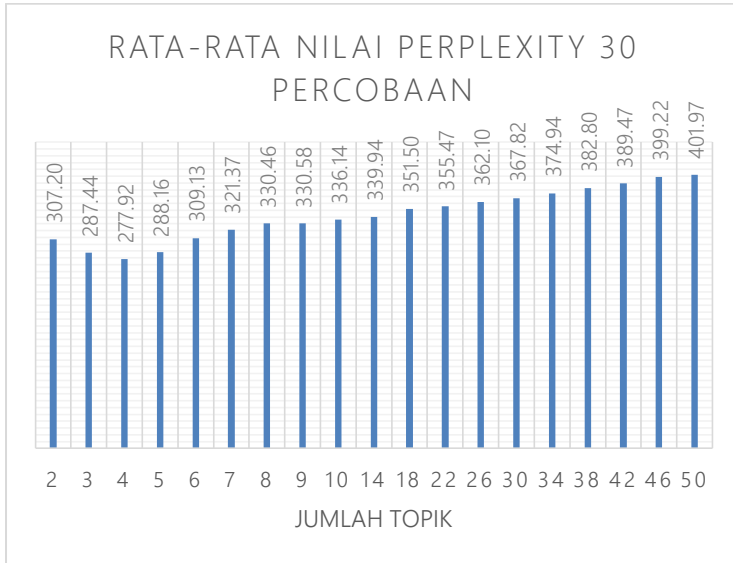
**Diagram 6.3 Standar Deviasi Nilai *Perplexity* 30 Percobaan**

Berdasarkan Diagram 6.3, diketahui bahwa standar deviasi nilai *perplexity* 4 topik untuk 30 kali percobaan adalah 2.54. jika dibandingkan dengan nilai rata-rata standar deviasi keseluruhan yaitu 2.81, angka 2.54 dapat dikatakan cukup stabil

#### **6.4.2.2. Penentuan Jumlah Topik tanpa *Stemming***

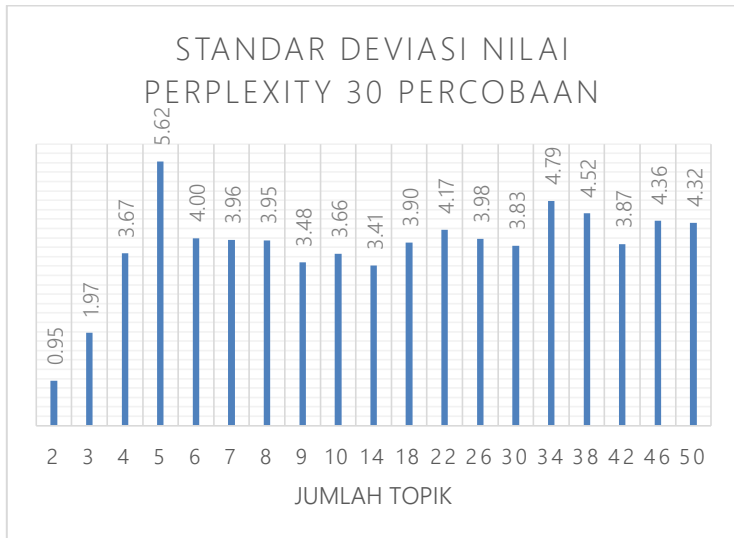
Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik tanpa *Stemming* dapat dilihat pada lampiran A3

#### 6.4.2.2.1. Analisis Hasil Eksperimen Nilai *Perplexity* untuk Penentuan Jumlah Topik



**Diagram 6.4 Rata-rata nilai *Perplexity* 30 percobaan**

Berdasarkan Diagram 6.4, nilai *perplexity* terendah terdapat pada jumlah topik 4 yaitu 277.92, dan tren nilai *perplexity* meningkat untuk jumlah topik yang semakin tinggi, sehingga 4 topik merupakan jumlah topik terbaik berdasarkan analisis nilai *perplexity*



**Diagram 6.5 Standar Deviasi Nilai *Perplexity* 30 Percobaan**

Berdasarkan Diagram 6.5, diketahui bahwa standar deviasi nilai *perplexity* 4 topik untuk 30 kali percobaan adalah 3.67. Jika dibandingkan dengan nilai rata-rata standar deviasi keseluruhan yaitu 3.81, angka 3.67 dapat dikatakan cukup stabil

#### **6.4.3. Jumlah Distribusi Dokumen Per Topik**

Distribusi dokumen per topik dilakukan sebagai bentuk analisis terhadap dominansi antara satu topik dengan topik yang lain dengan parameter jumlah dokumen per topik. Analisis distribusi dokumen dilakukan dengan melakukan klasifikasi dokumen mentah ke dalam model yang telah dibentuk. Peran analisis distribusi dokumen per topik dalam tahap validasi adalah untuk melihat apakah persebaran dokumen terpusat pada salah satu dari topik yang dibentuk atau tersebar secara merata. Analisis distribusi dokumen per topik dilakukan untuk topik yang dihasilkan melalui tahap *Stemming* dan yang tidak melalui tahap *Stemming*



#### 6.4.3.1. Distribusi Dokumen Per Topik dengan *Stemming*

Distribusi dokumen per topik dengan *Stemming* ditampilkan dalam Tabel 6.7

**Tabel 6.7 Distribusi Dokumen Per Topik dengan *Stemming***

Topik	Jumlah Distribusi Dokumen
Topik #0	10237
Topik #1	5884
Topik #2	6533
Topik #3	14590
Total dokumen dengan distribusi topik merata	30

Berdasarkan Tabel 6.7, dapat terlihat bahwa distribusi dokumen mayoritas berada pada topik #3, diikuti dengan topik #2 dan terdapat 30 dokumen yang tidak tergolong ke dalam topik manapun

#### 6.4.3.2. Distribusi Dokumen Per Topik tanpa *Stemming*

Distribusi dokumen per topik tanpa *Stemming* ditampilkan dalam Tabel 6.8

**Tabel 6.8 Distribusi Dokumen Per Topik tanpa *Stemming***

Topik	Jumlah Distribusi Dokumen
Topik #0	10437
Topik #1	15904
Topik #2	5579
Topik #3	5324
Total dokumen dengan distribusi topik merata	30

Berdasarkan Tabel 6.8, dapat terlihat bahwa distribusi dokumen mayoritas berada pada topik #1, diikuti dengan topik #0 dan terdapat 30 dokumen yang tidak tergolong ke dalam topik manapun

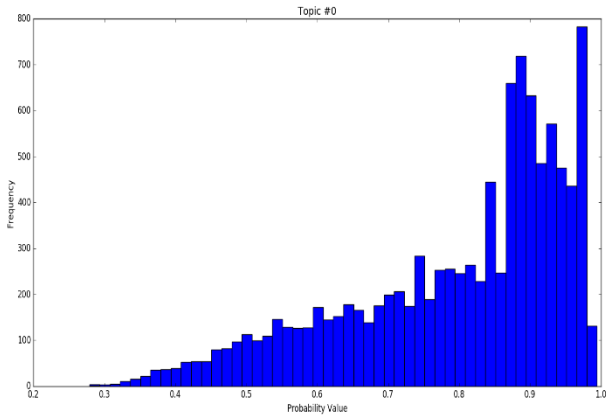
#### **6.4.4. Visualisasi Distribusi Probabilitas Dokumen per Topik**

Analisis distribusi dokumen per topik dilakukan sebagai bentuk analisis persebaran dokumen per topik dengan tujuan untuk mengetahui apakah terdapat dominansi pada satu topik terhadap topik lainnya. Namun analisis ini perlu dilakukan analisis lanjutan untuk mengetahui tingkat keyakinan suatu dokumen terhadap topik yang menjadi golongannya melalui probabilitas dokumen terhadap topiknya. Untuk itu perlu dilakukan analisis distribusi probabilitas dokumen per topik yang divisualisasikan dalam bentuk histogram agar lebih mudah dalam melakukan pengamatan. Analisis distribusi probabilitas dokumen per topik dilakukan untuk topik yang dihasilkan oleh model yang melalui tahap *Stemming* dan yang tidak melalui tahap *Stemming*

##### **6.4.4.1. Distribusi Probabilitas Dokumen per Topik dengan *Stemming***

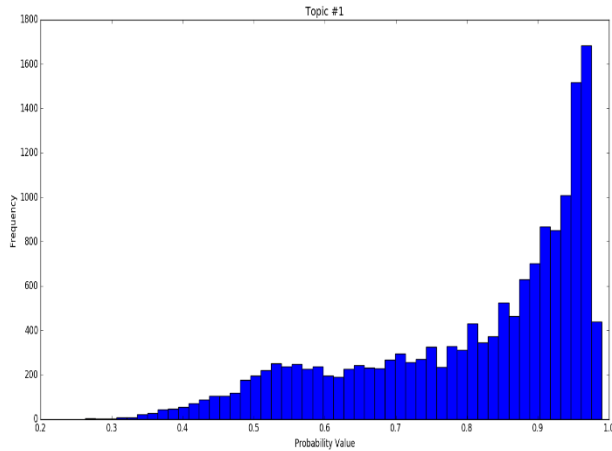
Berikut merupakan histogram distribusi probabilitas dokumen per topik dengan *Stemming*

#### 6.4.4.1.1. Histogram



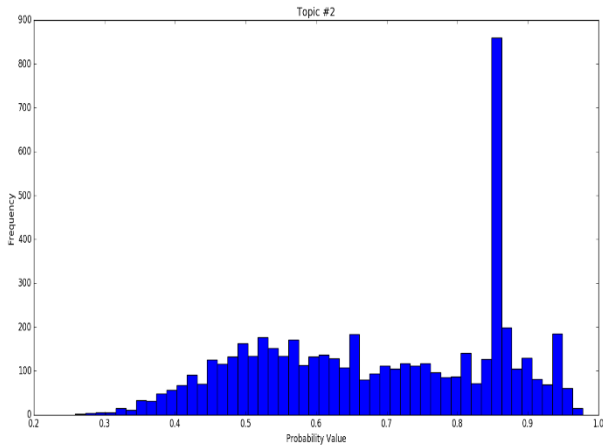
**Diagram 6.6 Histogram distribusi probabilitas dokumen pada Topik #0**

Berdasarkan Diagram 6.6, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.6 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.85 hingga 0.99



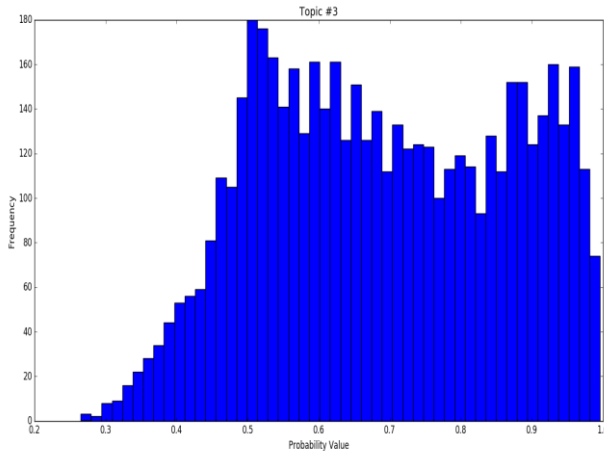
**Diagram 6.7 Histogram distribusi probabilitas dokumen pada Topik #1**

Berdasarkan Diagram 6.7, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.9 hingga 0.99



**Diagram 6.8 Histogram distribusi probabilitas dokumen pada Topik #2**

Berdasarkan Diagram 6.8, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.85 hingga 0.87



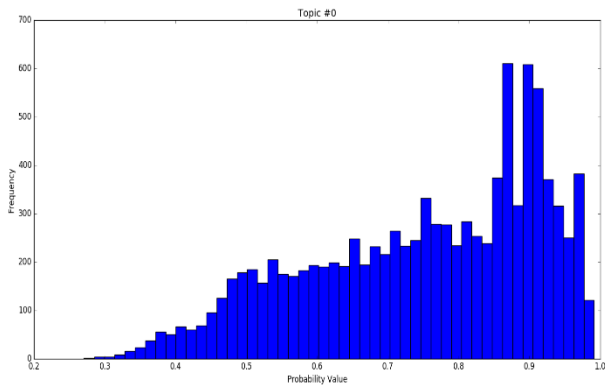
**Diagram 6.9 Histogram distribusi probabilitas dokumen pada Topik #3**

Berdasarkan Diagram 6.9, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.5 hingga 0.99

#### **6.4.4.2. Distribusi Probabilitas Topik Dokumen per Topik dengan *Stemming***

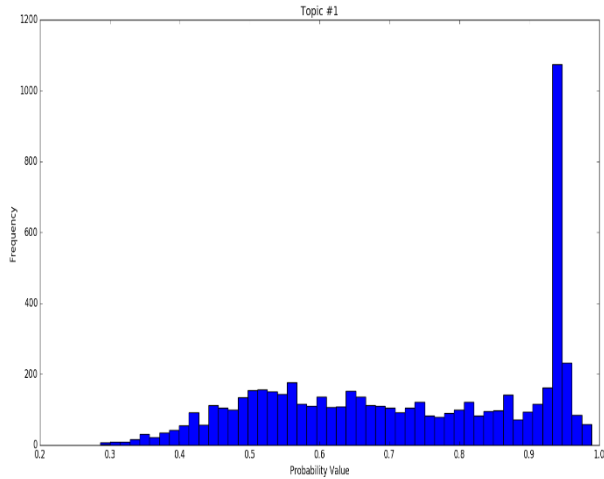
Berikut merupakan histogram distribusi probabilitas dokumen per topik dengan *Stemming*

### 6.4.4.2.1. Histogram



**Diagram 6.10 Histogram distribusi probabilitas dokumen pada Topik #0**

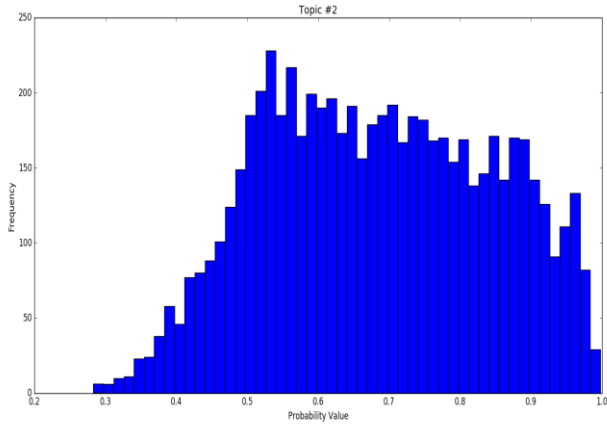
Berdasarkan Diagram 6.10, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.85 hingga 0.95



**Diagram 6.11 Histogram distribusi probabilitas dokumen pada Topik #1**

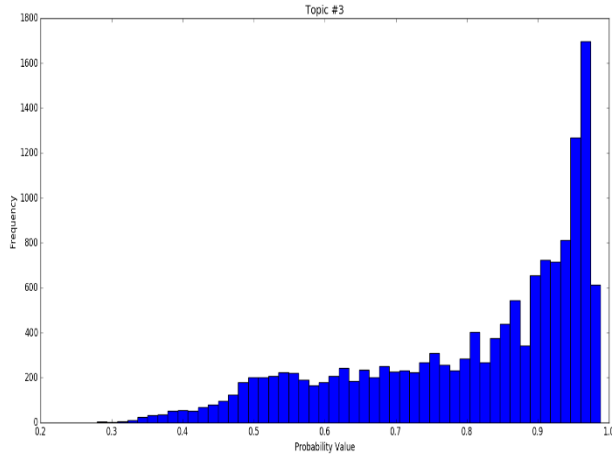
Berdasarkan Diagram 6.11, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.93 hingga 0.96





**Diagram 6.12 Histogram distribusi probabilitas dokumen pada Topik #2**

Berdasarkan Diagram 6.12, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.5 hingga 0.9



**Diagram 6.13 Histogram distribusi probabilitas dokumen pada Topik #3**

Berdasarkan Diagram 6.13, distribusi probabilitas dinilai meyakinkan dengan mayoritas dokumen memiliki probabilitas topik lebih dari 0.5 dan tingginya jumlah dokumen yang memiliki probabilitas dalam rentang 0.9 hingga 0.99

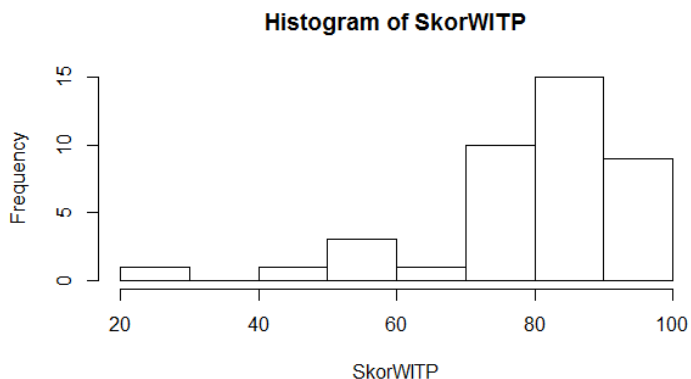
## **6.5. Hasil Uji Koherensi Topik**

Penyebaran kuesioner uji koherensi topik dilakukan sebanyak dua kali. Penyebaran pertama dilakukan pada tanggal 10 - 14 Mei 2017 dengan total 102 responden dan penyebaran kedua dilakukan pada tanggal 2 – 4 Juni 2017 dengan total 122 responden. Hasil rekapitulasi tanggapan dari kuesioner uji koherensi topik dapat dilihat pada lampiran A4 dan A5

### **6.5.1. Histogram**

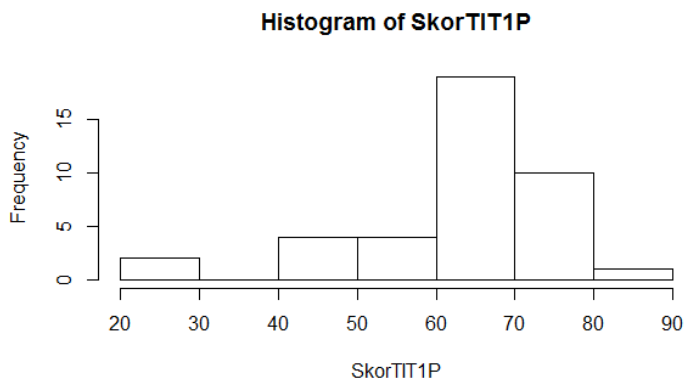
#### **6.5.1.1. Berbasis Pertanyaan**

Analisis ini bertujuan untuk mengetahui tingkat kesulitan dari pertanyaan yang diajukan secara keseluruhan. Tingkat kesulitan pertanyaan diukur dengan persentase ketepatan responden dalam menjawab pertanyaan sesuai *task* pada uji koherensi topik. Berikut merupakan histogram rekapitulasi hasil uji koherensi topik yang dirangkum berbasis skor dari setiap pertanyaan.



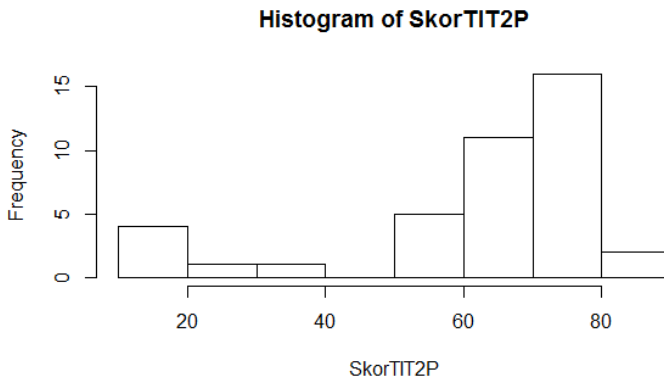
**Diagram 6.14 Histogram skor *Word Intrusion task* berbasis pertanyaan**

Diagram 6.14 menampilkan distribusi skor untuk pertanyaan kuesioner yang tergolong dalam *Word Intrusion Task*. Berdasarkan distribusi yang ditampilkan dalam histogram, mayoritas persentase ketepatan untuk setiap pertanyaan berada pada rentang 70% - 100%, sehingga pertanyaan-pertanyaan yang tergolong dalam kategori *Word Intrusion task* dapat dikatakan sangat mudah



**Diagram 6.15 Histogram skor *Topic Intrusion task 1* berbasis pertanyaan**

Diagram 6.15 menampilkan distribusi skor untuk pertanyaan kuesioner yang tergolong dalam *Topic Intrusion task 1*. Berdasarkan distribusi yang ditampilkan dalam histogram, mayoritas persentase ketepatan untuk setiap pertanyaan berada pada rentang 60% - 80%, sehingga pertanyaan-pertanyaan yang tergolong dalam kategori *Topic Intrusion task 1* dapat dikatakan mudah



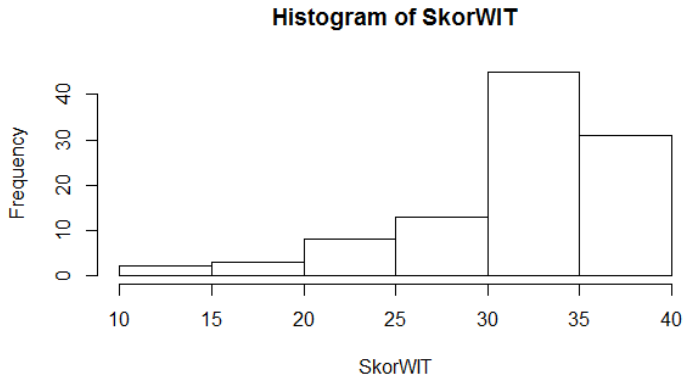
**Diagram 6.16 Histogram skor *Topic Intrusion task 2* berbasis pertanyaan**

Diagram 6.16 menampilkan distribusi skor untuk pertanyaan kuesioner yang tergolong dalam *Topic Intrusion task 2*. Berdasarkan distribusi yang ditampilkan dalam histogram, mayoritas persentase ketepatan untuk setiap pertanyaan berada pada rentang 60% - 80%, sehingga pertanyaan-pertanyaan yang tergolong dalam kategori *Topic Intrusion task 2* dapat dikatakan mudah

#### **6.5.1.2. Berbasis Responden**

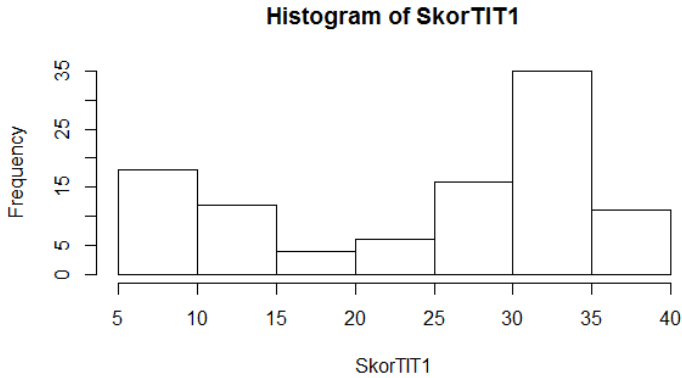
Analisis ini bertujuan untuk mengetahui tingkat interpretasi responden terhadap topik melalui pertanyaan yang diajukan secara keseluruhan. tingkat interpretasi responden diukur dengan jumlah pertanyaan yang dijawab secara tepat untuk masing-masing *task* pada uji koherensi topik. Untuk

setiap *task* dalam uji koherensi topik, terdapat 40 pertanyaan yang diajukan, sehingga skor maksimal yang dapat diperoleh oleh responden adalah 40. Berikut merupakan histogram rekapitulasi hasil uji koherensi topik yang dirangkum berbasis responden.



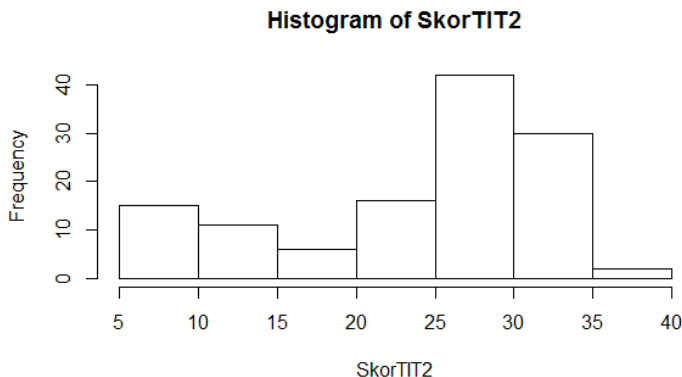
**Diagram 6.17 Histogram skor *Word Intrusion task* berbasis responden**

Diagram 6.17 menampilkan distribusi skor ketepatan responden dalam menjawab pertanyaan yang tergolong dalam *Word Intrusion Task*. Berdasarkan distribusi yang ditampilkan dalam histogram, sekitar 75 responden menjawab secara tepat 30-40 pertanyaan, sehingga tingkat interpretasi responden terhadap topik terhadap pertanyaan-pertanyaan yang tergolong dalam kategori *Word Intrusion task* dapat dikatakan sangat baik



**Diagram 6.18 Histogram skor *Topic Intrusion task 1* berbasis responden**

Diagram 6.18 menampilkan distribusi skor ketepatan responden dalam menjawab pertanyaan yang tergolong dalam *Topic Intrusion task 1*. Berdasarkan distribusi yang ditampilkan dalam histogram, sekitar 35 responden menjawab secara tepat 30-35 pertanyaan, diikuti sekitar 17 responden menjawab secara tepat 5-10 pertanyaan, sehingga tingkat interpretasi responden terhadap topik terhadap pertanyaan-pertanyaan yang tergolong dalam kategori *Topic Intrusion task 1* dapat dikatakan baik.



**Diagram 6.19 Histogram skor *Topic Intrusion task 2* berbasis responden**

Diagram 6.19 menampilkan distribusi skor ketepatan responden dalam menjawab pertanyaan yang tergolong dalam *Topic Intrusion task 2*. Berdasarkan distribusi yang ditampilkan dalam histogram, sekitar 75 responden menjawab secara tepat 25-35 pertanyaan, sehingga tingkat interpretasi responden terhadap topik terhadap pertanyaan-pertanyaan yang tergolong dalam kategori *Topic Intrusion task 1* dapat dikatakan baik.

### 6.5.2. Uji Hipotesis

Uji Hipotesis dilakukan dengan tujuan untuk membuktikan hipotesis (dugaan sementara) berdasarkan metode statistika. Dalam melakukan uji hipotesis, tahap-tahap yang dilakukan adalah mendefinisikan pernyataan hipotesis, melakukan uji *variance*, melakukan uji *means* kemudian menyusun kesimpulan berdasarkan kedua uji tersebut. Kesimpulan uji *variance* dan uji *means* dilakukan dengan membandingkan *p-value* dengan tingkat signifikansi. Tingkat signifikansi ( $\alpha$ ) merupakan probabilitas penolakan hipotesis nol ketika hipotesis tersebut benar. Tingkat signifikansi akan menjadi pembanding terhadap nilai *p-value* untuk menentukan



apakah hipotesis nol diterima atau ditolak, hipotesis nol ditolak jika nilai  $p$ -value kurang dari nilai  $\alpha$ . Pada kasus pengujian hipotesis ini dipilih tingkat signifikansi  $\alpha = 0.05$ . Untuk mempertajam analisis, uji ANOVA dibagi berdasarkan metode uji koherensi topik dengan dibantu *tools* Minitab 17. Untuk melakukan uji hipotesis, diperlukan data yang terdapat pada Tabel 6.9

**Tabel 6.9 Tabel Pendahuluan Uji Hipotesis**

	Rata-rata		St Dev	
	#	%	#	%
WIT	31.96	79.90	5.57	14.00
WIT -Stem	15.66	78.28	2.67	13.33
WIT +Stem	16.30	81.52	3.99	19.96
TIT 1	24.86	62.16	10.31	25.78
TIT 1 -Stem	12.28	61.42	4.44	22.19
TIT 1 +Stem	13.26	66.32	4.67	23.35
TIT 2	24.32	60.80	8.33	20.83
TIT 1 -Stem	12.13	60.66	5.07	25.35
TIT 1 +Stem	12.88	64.39	5.93	29.66

#### **6.5.2.1. Uji Hipotesis *Word Intrusion task* dan *Topic Intrusion task 1***

##### **Pernyataan Hipotesis *Word Intrusion task* dan *Topic Intrusion task 1***

$H_0$ :  $\sigma$  *Word Intrusion task* =  $\sigma$  *Topic Intrusion task 1*

$H_1$ :  $\sigma$  *Word Intrusion task*  $\neq$   $\sigma$  *Topic Intrusion task 1*

##### **Uji Variance**

### Test and CI for Two Variances: WIT, TIT 1

#### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
WIT	102	5.599	31.345	(4.549, 7.025)
TIT 1	102	10.310	106.298	(9.531, 11.371)

Ratio of standard deviations = 0.543

Ratio of variances = 0.295

#### Tests

Method	DF1	DF2	Test	
			Statistic	P-Value
Bonett	1	—	43.73	0.000
Levene	1	202	27.99	0.000

**Gambar 6.1 Uji variance WIT dan TIT 1**

Hasil pengujian yang ditampilkan pada Gambar 6.1 menunjukkan bahwa *p-value* dengan metode *Bonett* dan *Levene* senilai 0.000, yang mana nilai tersebut kurang dari nilai  $\alpha$ , sehingga dapat dikatakan bahwa nilai *variance* dari *Word Intrusion task* dan *Topic Intrusion task 1* dinyatakan berbeda

### Uji Means

### Two-Sample T-Test and CI: WIT, TIT 1

Two-sample T for WIT vs TIT 1

	N	Mean	StDev	SE Mean
WIT	102	31.96	5.60	0.55
TIT 1	102	24.9	10.3	1.0

Difference =  $\mu$  (WIT) -  $\mu$  (TIT 1)

Estimate for difference: 7.10

95% CI for difference: (4.80, 9.39)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = 6.11 P-Value = 0.000 DF = 155

**Gambar 6.2 Uji means WIT dan TIT 1**

Hasil pengujian yang ditampilkan pada Gambar 6.2 menunjukkan bahwa *p-value* senilai 0.000, yang mana nilai tersebut lebih kecil dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol ditolak

Berdasarkan uji hipotesis *Word Intrusion task* dan *Topic Intrusion task 1*, disimpulkan bahwa kedua sampel memiliki *variance* dan *Means* yang berbeda signifikan, sehingga diperlukan analisis lebih lanjut terhadap kesimpulan ini. Terdapat banyak kemungkinan rendahnya nilai TIT 1 jika dibandingkan dengan WIT, dalam kasus ini, hal yang paling disoroti adalah faktor kelelahan responden dan responden kesulitan dalam mengisi kuesioner akibat tampilan visual yang kurang baik.

Untuk menguji kedua asumsi yang muncul, dirasa perlu melakukan beberapa tindakan. Pertama, penyebaran ulang kuesioner khusus bagian *Topic Intrusion task* untuk mengantisipasi faktor kelelahan responden sehingga didapatkan tingkat fokus yang sama seperti WIT pada penyebaran kuesioner tahap pertama. Kedua, memperbaiki tampilan visual kuesioner untuk menjaga agar tidak terlalu banyak informasi yang ditampilkan dalam satu pertanyaan yang dapat membingungkan responden. Perbaikan tampilan visual dilakukan dengan membatasi pilihan jawaban yang sebelumnya terdapat 20 kata dalam satu topik menjadi 7 kata agar tetap

dalam satu baris sesuai dengan perencanaan materi kuesioner yang terdapat pada sub bab 4.3.5.2.2.

#### 6.5.2.2. Uji Hipotesis *Topic Intrusion task 1* dan *Topic Intrusion task 2*

##### Pernyataan Hipotesis *Topic Intrusion task 1* dan *Topic Intrusion task 2*

$H_0: \sigma \text{ Topic Intrusion task 1} = \sigma \text{ Topic Intrusion task 2}$

$H_1: \sigma \text{ Topic Intrusion task 1} \neq \sigma \text{ Topic Intrusion task 2}$

#### Uji Variance

##### Test and CI for Two Variances: TIT 1, TIT 2

###### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
TIT 1	102	10.310	106.298	(9.531, 11.371)
TIT 2	122	8.333	69.442	(7.485, 9.429)

Ratio of standard deviations = 1.237

Ratio of variances = 1.531

###### Tests

Method	DF1	DF2	Test	
			Statistic	P-Value
Bonett	—	—	—	0.003
Levene	1	222	5.57	0.019

Gambar 6.3 Uji variance TIT 1 dan TIT 2

Hasil pengujian yang ditampilkan pada Gambar 6.3 menunjukkan bahwa *p-value* dengan metode *Bonett* senilai 0.003 dan metode *Levene* senilai 0.019, yang mana kedua nilai tersebut kurang dari nilai  $\alpha$ , sehingga dapat dikatakan bahwa

nilai *variance* dari *Topic Intrusion task 1* dan *Topic Intrusion task 2* dinyatakan berbeda

### Uji Means

#### Two-Sample T-Test and CI: TIT 1, TIT 2

Two-sample T for TIT 1 vs TIT 2

	N	Mean	StDev	SE Mean
TIT 1	102	24.9	10.3	1.0
TIT 2	122	24.32	8.33	0.75

Difference =  $\mu$  (TIT 1) -  $\mu$  (TIT 2)

Estimate for difference: 0.54

95% CI for difference: (-1.96, 3.05)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = 0.43 P-Value = 0.669 DF = 193

**Gambar 6.4 Uji Means TIT 1 dan TIT 2**

Hasil pengujian yang ditampilkan pada Gambar 6.4 menunjukkan bahwa *p-value* senilai 0.669, yang mana nilai tersebut lebih besar dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa tidak terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol gagal tolak.

Berdasarkan uji hipotesis *Topic Intrusion task 1* dan *Topik Intusion task 2*, disimpulkan bahwa kedua sampel memiliki *variance* yang berbeda namun *Means* yang sama, sehingga hasil ini dapat memberikan gambaran terhadap kedua asumsi yang muncul pada analisis lanjutan uji hipotesis *Word Intrusion task* dan *Topik Intusion task 1*.

Asumsi adanya kelelahan responden dan adanya kesulitan responden terhadap tampilan kuesioner bukan merupakan faktor yang berpengaruh terhadap kedua sampel, hal ini dibuktikan dengan uji *Means* yang menunjukkan bahwa  $H_0$  gagal tolak, atau dalam artian lain *Means* untuk kedua sampel dinyatakan sama.

### 6.5.2.3. Uji Hipotesis *Word Intrusion task* dan *Topic Intrusion task 2*

#### Pernyataan Hipotesis *Word Intrusion task* dan *Topic Intrusion task 2*

$H_0: \sigma \text{ Word Intrusion task} = \sigma \text{ Topic Intrusion task 2}$

$H_1: \sigma \text{ Word Intrusion task} \neq \sigma \text{ Topic Intrusion task 2}$

#### Uji Variance

##### Test and CI for Two Variances: WIT, TIT 2

###### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
WIT	102	5.599	31.345	(4.549, 7.025)
TIT 2	122	8.333	69.442	(7.485, 9.429)

Ratio of standard deviations = 0.672

Ratio of variances = 0.451

###### Tests

Method	DF1	DF2	Test	
			Statistic	P-Value
Bonett	—	—	—	0.000
Levene	1	222	11.91	0.001

Gambar 6.5 Uji variance WIT dan TIT 2

Hasil pengujian yang ditampilkan pada Gambar 6.5 menunjukkan bahwa *p-value* dengan metode *Bonett* senilai 0.000 dan metode *Levene* senilai 0.001, yang mana kedua nilai tersebut kurang dari nilai  $\alpha$ , sehingga dapat dikatakan bahwa nilai *variance* dari *Word Intrusion task* dan *Topic Intrusion task 1* dinyatakan berbeda

## Uji Means

### Two-Sample T-Test and CI: WIT, TIT 2

Two-sample T for WIT vs TIT 2

	N	Mean	StDev	SE Mean
WIT	102	31.96	5.60	0.55
TIT 2	122	24.32	8.33	0.75

Difference =  $\mu$  (WIT) -  $\mu$  (TIT 2)

Estimate for difference: 7.641

95% CI for difference: (5.796, 9.487)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = 8.16 P-Value = 0.000 DF = 212

**Gambar 6.6 Uji Means WIT 1 dan TIT 2**

Hasil pengujian yang ditampilkan pada Gambar 6.6 menunjukkan bahwa *p-value* senilai 0.000, yang mana nilai tersebut lebih kecil dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol ditolak.

Berdasarkan seluruh uji hipotesis yang telah dilakukan, dapat disimpulkan bahwa metode *Word Intrusion task* secara inheren lebih mudah diinterpretasi oleh manusia jika dibandingkan dengan metode *Topic Intrusion task* baik TIT 1 maupun TIT 2.

#### 6.5.2.4. Uji Hipotesis Pengaruh *Stemming* pada *Word Intrusion Task*

##### Pernyataan Hipotesis Pengaruh *Stemming* pada *Word Intrusion task*

$H_0$ :  $\sigma$  WIT tanpa *Stem* =  $\sigma$  WIT dengan *Stem*

$H_1$ :  $\sigma$  WIT tanpa *Stem*  $\neq$   $\sigma$  WIT dengan *Stem*

##### Uji *Variance*

###### Test and CI for Two Variances: WIT -Stem, WIT +Stem

###### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
WIT -Stem	102	5.096	25.968	(4.681, 5.656)
WIT +Stem	102	5.962	35.543	(5.482, 6.610)

Ratio of standard deviations = 0.855

Ratio of variances = 0.731

###### Tests

Method	DF1	DF2	Test	
			Statistic	P-Value
Bonett	1	—	5.31	0.021
Levene	1	202	1.47	0.227

Gambar 6.7 Uji *variance* WIT -Stem dan WIT +Stem

Hasil pengujian yang ditampilkan pada Gambar 6.7 menunjukkan bahwa *p-value* dengan metode *Bonett* senilai 0.021 dan metode *Levene* senilai 0.221, yang mana nilai dari metode *Bonett* tersebut kurang dari nilai  $\alpha$ , sementara nilai dari metode *Levene* lebih besar dari  $\alpha$ , namun rata-rata kedua nilai tersebut lebih besar dari  $\alpha$ , sehingga dapat dikatakan bahwa nilai *variance* dari *Word Intrusion task* tanpa *Stem* dan *Word Intrusion task* dengan *Stem* dinyatakan sama



## Uji Means

### Two-Sample T-Test and CI: WIT -Stem, WIT +Stem

Two-sample T for WIT -Stem vs WIT +Stem

	N	Mean	StDev	SE Mean
WIT -Stem	102	12.28	5.10	0.50
WIT +Stem	102	13.26	5.96	0.59

Difference =  $\mu$  (WIT -Stem) -  $\mu$  (WIT +Stem)

Estimate for difference: -0.980

95% CI for difference: (-2.512, 0.551)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = -1.26 P-Value = 0.208 DF = 197

**Gambar 6.8 Uji Means WIT -Stem dan WIT +Stem**

Hasil pengujian yang ditampilkan pada Gambar 6.8 menunjukkan bahwa *p-value* senilai 0.208, yang mana nilai tersebut lebih besar dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa tidak terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol gagal tolak

### 6.5.2.5. Uji Hipotesis Pengaruh *Stemming* pada *Topic Intrusion task 1*

#### Pernyataan Hipotesis Pengaruh *Stemming* pada *Topic Intrusion task 1*

$H_0: \sigma \text{ TIT 1 tanpa Stem} = \sigma \text{ TIT 1 dengan Stem}$

$H_1: \sigma \text{ TIT 1 tanpa Stem} \neq \sigma \text{ TIT 1 dengan Stem}$

#### Uji Variance

##### Test and CI for Two Variances: TIT 1 -Stem, TIT 1 +Stem

###### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
TIT 1 -Stem	102	2.679	7.178	(2.285, 3.203)
TIT 1 +Stem	102	4.012	16.095	(3.132, 5.240)

Ratio of standard deviations = 0.668

Ratio of variances = 0.446

###### Tests

Method	DF1	DF2	Statistic	P-Value
Bonett	1	—	4.64	0.031
Levene	1	202	2.36	0.126

Gambar 6.9 Uji variance TIT 1 -Stem dan TIT 1 +Stem

Hasil pengujian yang ditampilkan pada Gambar 6.9 menunjukkan bahwa *p-value* dengan metode *Bonett* senilai 0.031 dan metode *Levene* senilai 0.126, yang mana nilai dari metode *Bonett* tersebut kurang dari nilai  $\alpha$ , sementara nilai dari metode *Levene* lebih besar dari  $\alpha$ , namun rata-rata kedua nilai tersebut lebih besar dari  $\alpha$ , sehingga dapat dikatakan bahwa nilai *variance* dari *Topic Intrusion task 1* tanpa *Stem* dan *Topic Intrusion task 1* dengan *Stem* dinyatakan sama

## Uji Means

### Two-Sample T-Test and CI: TIT 1 -Stem, TIT 1 +Stem

Two-sample T for TIT 1 -Stem vs TIT 1 +Stem

	N	Mean	StDev	SE Mean
TIT 1 -Stem	102	15.66	2.68	0.27
TIT 1 +Stem	102	16.30	4.01	0.40

Difference =  $\mu$  (TIT 1 -Stem) -  $\mu$  (TIT 1 +Stem)

Estimate for difference: -0.647

95% CI for difference: (-1.590, 0.296)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = -1.35 P-Value = 0.177 DF = 176

**Gambar 6.10 Uji Means TIT 1 -Stem dan TIT 1 +Stem**

Hasil pengujian yang ditampilkan pada Gambar 6.10 menunjukkan bahwa *p-value* senilai 0.177, yang mana nilai tersebut lebih besar dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa tidak terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol gagal tolak

### 6.5.2.6. Uji Hipotesis Pengaruh *Stemming* pada *Topic Intrusion task 2*

#### Pernyataan Hipotesis Pengaruh *Stemming* pada *Topic Intrusion task 1*

$H_0$ :  $\sigma$  TIT 1 tanpa *Stem* =  $\sigma$  TIT 1 dengan *Stem*

$H_1$ :  $\sigma$  TIT 1 tanpa *Stem*  $\neq$   $\sigma$  TIT 1 dengan *Stem*

#### Uji *Variance*

##### Test and CI for Two Variances: TIT 2 - Stem, TIT 2 +Stem

###### Statistics

Variable	N	StDev	Variance	95% CI for StDevs
TIT 2 - Stem	122	4.457	19.867	(3.981, 5.071)
TIT 2 +Stem	122	4.690	21.993	(4.202, 5.319)

Ratio of standard deviations = 0.950

Ratio of variances = 0.903

###### Tests

Method	DF1	DF2	Test	
			Statistic	P-Value
Bonett	1	—	0.36	0.549
Levene	1	242	0.64	0.424

Gambar 6.11 Uji *variance* TIT 2 -Stem dan TIT 2 +Stem

Hasil pengujian yang ditampilkan pada Gambar 6.11 menunjukkan bahwa *p-value* dengan metode *Bonett* senilai 0.549 dan metode *Levene* senilai 0.429, yang mana kedua nilai tersebut lebih besar dari nilai  $\alpha$ , sehingga dapat dikatakan bahwa nilai *variance* dari *Topic Intrusion task 2* tanpa *Stem* dan *Topic Intrusion task 2* dengan *Stem* dinyatakan sama

## Uji Means

### Two-Sample T-Test and CI: TIT 2 - Stem, TIT 2 +Stem

Two-sample T for TIT 2 - Stem vs TIT 2 +Stem

	N	Mean	StDev	SE Mean
TIT 2 - Stem	122	12.13	4.46	0.40
TIT 2 +Stem	122	12.88	4.69	0.42

Difference =  $\mu$  (TIT 2 - Stem) -  $\mu$  (TIT 2 +Stem)

Estimate for difference: -0.746

95% CI for difference: (-1.900, 0.408)

T-Test of difference = 0 (vs  $\neq$ ): T-Value = -1.27 P-Value = 0.204 DF = 241

**Gambar 6.12 Uji Means TIT 2 -Stem dan TIT 2 +Stem**

Hasil pengujian yang ditampilkan pada Gambar 6.12 menunjukkan bahwa *p-value* senilai 0.204, yang mana nilai tersebut lebih besar dari nilai  $\alpha$ , nilai tersebut menunjukkan bahwa tidak terdapat nilai *Means* yang signifikan antara kedua sampel, sehingga hipotesis nol gagal tolak

### 6.5.2.7. Kesimpulan Uji Hipotesis

Berdasarkan hasil uji hipotesis yang telah dilakukan, berikut merupakan rekapitulasi uji *variance* yang ditunjukkan pada Tabel 6.10 dan rekapitulasi uji *Means* yang ditunjukkan pada Tabel 6.11

**Tabel 6.10 Tabel Rekapitulasi Uji Variance**

Uji Variance ( $\alpha = 0.05$ )			
Sampel	<i>p-value</i>		kesimpulan
	<i>Bonett</i>	<i>Lavene</i>	
WIT -Stem & WIT +Stem	0.021	0.227	<i>variance</i> sama
TIT 1 -Stem & TIT 1 +Stem	0.031	0.126	<i>variance</i> sama
TIT 2 -Stem & TIT 2 +Stem	0.549	0.424	<i>variance</i> sama
WIT & TIT 1	0.000	0.000	<i>variance</i> beda
WIT & TIT 2	0.000	0.001	<i>variance</i> beda
TIT 1 & TIT 2	0.003	0.019	<i>variance</i> beda

**Tabel 6.11 Tabel Rekapitulasi Uji Means**

Uji Means		
a = 0.05	p-value	kesimpulan
WIT -Stem vs WIT +Stem	0.208	Means sama, H <sub>0</sub> gagal tolak
TIT 1 -Stem vs TIT 1 +Stem	0.177	Means sama, H <sub>0</sub> gagal tolak
TIT 2 -Stem vs TIT 2 +Stem	0.204	Means sama, H <sub>0</sub> gagal tolak
WIT vs TIT 1	0.000	Means beda, H <sub>0</sub> ditolak
WIT vs TIT 2	0.000	Means beda, H <sub>0</sub> ditolak
TIT 1 vs TIT 2	0.669	Means sama, H <sub>0</sub> gagal tolak

Berdasarkan uji hipotesis yang dilakukan melalui uji *variance* dan uji *Means*, terdapat dua kesimpulan yang dapat diambil. Pertama, nilai rata-rata metode WIT lebih tinggi dan berbeda signifikan jika dibandingkan dengan nilai rata-rata metode TIT 1 dan TIT 2, hal ini menunjukkan bahwa metode *Word Intrusion task* secara inheren mampu diinterpretasi lebih baik oleh responden. Hal ini dibuktikan pula dengan nilai rata-rata metode TIT 2 yang tidak signifikan terhadap nilai rata-rata TIT 1 meskipun dalam pelaksanaan TIT 2 telah dilakukan perbaikan seperlunya dan dilakukan secara terpisah sebagai usaha untuk mengeliminasi asumsi penyebab dari skor pada TIT 1 yang lebih rendah dari WIT. Kedua, nilai rata-rata metode WIT, TIT 1, dan TIT 2 dengan *Stem* memberikan hasil yang lebih baik. Hal ini menunjukkan bahwa proses *Stemming* berpengaruh positif terhadap pemodelan yang dilakukan dalam kasus ini, namun berdasarkan uji *variance* dan uji *Means*, perbedaan ini tidak signifikan.

### 6.5.3. Uji ANOVA (Analysis of Variance)

Uji ANOVA bertujuan untuk untuk membuktikan hipotesis (dugaan sementara) berdasarkan metode statistika, biasanya digunakan untuk melihat signifikansi pada rata-rata

dari dua atau lebih grup yang berbeda. Uji ANOVA dalam kasus ini bertujuan untuk menguji signifikansi rata-rata pada setiap topik, sehingga dapat diketahui topik mana yang berbeda secara signifikan dari topik yang lain. Untuk mempertajam analisis, uji ANOVA dibagi berdasarkan metode uji koherensi topik dengan dibantu *tools* Minitab 17.

Dalam melakukan uji ANOVA, tahap-tahap yang dilakukan adalah mendefinisikan pernyataan hipotesis, melakukan uji ANOVA kemudian menyusun kesimpulan berdasarkan kedua uji tersebut. Kesimpulan ANOVA dilakukan dengan membandingkan *p-value* dengan tingkat signifikansi. Tingkat signifikansi ( $\alpha$ ) merupakan probabilitas penolakan hipotesis nol ketika hipotesis tersebut benar. Tingkat signifikansi akan menjadi pembanding terhadap nilai *p-value* untuk menentukan apakah hipotesis nol diterima atau ditolak, hipotesis nol ditolak jika nilai *p-value* kurang dari nilai  $\alpha$ . Pada kasus pengujian hipotesis ini dipilih tingkat signifikansi  $\alpha = 0.05$ .

Sebelum melakukan uji anova, terlebih dahulu data kelompok topik dilakukan uji *variance*. Uji *variance* dilakukan untuk setiap kelompok topik pada setiap metode uji koherensi topik dengan hasil yang ditampilkan pada Tabel 6.12

**Tabel 6.12 Tabel Hasil Uji *Variance***

	<i>Variance p-value</i>		Kesimpulan
	Multiple Comparisons	Lavene	
WIT	0.644	0.518	<i>Variance</i> sama
TIT 1	0.059	0.235	<i>Variance</i> sama
TIT 2	0.000	0.004	<i>Variance</i> beda

Perbedaan varian ditunjukkan dengan melakukan perbandingan pada nilai *p-value* baik dengan metode multiple comparisons dan Lavene dengan nilai alpha ( $\alpha=0.05$ ), jika nilai

*p-value* lebih tinggi dari nilai  $\alpha$ , maka *variance* pada kedua kelompok uji koherensi topik dinyatakan berbeda signifikan.

Hasil uji *variance* menunjukkan *variance* pada topik topik dalam *Word Intrusion task* dan *Topic Intrusion task 1* memiliki *variance* yang sama, sementara topik topik dalam *Topic Intrusion task 2* memiliki *variance* berbeda. Topik-topik dalam uji koherensi yang memiliki *variance* yang sama akan dilakukan uji ANOVA dengan metode Tukey, sementara topik-topik dalam uji koherensi yang memiliki *variance* yang berbeda akan dilakukan uji anova dengan metode Games-Howell. Perbedaan metode ini akan tetap menghasilkan luaran yang serupa. Metode Tukey maupun Games Howell akan mengelompokkan nilai rata-rata ke dalam kelompok huruf. Suatu topik akan dinyatakan memiliki rata-rata yang sama jika digolongkan pada kelompok huruf yang sama, demikian pula sebaliknya, rata-rata topik akan dinyatakan berbeda signifikan jika digolongkan pada kelompok huruf yang berbeda.

### 6.5.3.1. Uji ANOVA *Word Intrusion Task*

#### Pernyataan hipotesis uji ANOVA *Word Intrusion task*

H0: semua *Means* untuk setiap sampel sama

H1: terdapat setidaknya satu *Means* yang berbeda

#### Uji Anova

Berikut merupakan informasi pengelompokan dengan Metode Tukey dengan Confidence Level 95% ( $\alpha = 0.05$ ) yang ditunjukkan pada Gambar 6.13

Tukey Pairwise Comparisons				
Grouping Information Using the Tukey Method and 95% Confidence				
Factor	N	Mean	Grouping	
WIT T1	102	8.559	A	
WIT T0	102	8.039	A B	
WIT T2	102	7.716	B	
WIT T3	102	7.647	B	
Means that do not share a letter are significantly different.				

**Gambar 6.13** Informasi pengelompokan sampel pada *Word Intrusion Task*



Untuk kasus uji ANOVA *Word Intrusion Task*, yang tergolong dalam kelompok huruf A adalah topik 1 dan topik 0, sementara yang tergolong dalam kelompok huruf B adalah topik 0, 2 dan topik 3. Topik 0 tergolong ke dalam kelompok huruf A dan B karena memiliki nilai *Means* yang sama dengan topik topik yang terdapat dalam kedua kelompok huruf

Games-Howell Simultaneous Tests for Differences of Means					
Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
TIT 2 T1 - TIT 2 T0	0.574	0.324	(-0.265, 1.413)	1.77	0.290
TIT 2 T2 - TIT 2 T0	-1.467	0.272	(-2.171, -0.763)	-5.40	0.000
TIT 2 T3 - TIT 2 T0	0.164	0.315	(-0.646, 0.974)	0.52	0.954
TIT 2 T2 - TIT 2 T1	-2.041	0.293	(-2.799, -1.283)	-6.96	0.000
TIT 2 T3 - TIT 2 T1	-0.410	0.334	(-1.267, 0.447)	-1.23	0.609
TIT 2 T3 - TIT 2 T2	1.631	0.283	( 0.898, 2.365)	5.75	0.000

**Gambar 6.14** Perbandingan berpasangan setiap sampel untuk *Word Intrusion Task*

Gambar 6.14 menunjukkan perbandingan berpasangan dari masing masing topik dalam uji ANOVA *Word Intrusion Task*. Berdasarkan Gambar 6.14 , perbedaan varian ditunjukkan dengan melakukan perbandingan pada nilai *p-value* dan nilai alpha ( $\alpha=0.05$ ), jika nilai *p-value* lebih tinggi dari nilai alpha, maka rata-rata kedua topik dinyatakan berbeda signifikan. Dengan demikian, perbandingan berpasangan untuk setiap topik telah menunjukkan kesesuaian dengan yang ditunjukkan dalam tabel informasi pengelompokan

### 6.5.3.2. Uji ANOVA *Topic Intrusion task 1*

#### Pernyataan hipotesis uji ANOVA *Topic Intrusion task 1*

H0: semua *Means* untuk setiap sampel sama

H1: terdapat setidaknya satu *Means* yang berbeda

#### Uji Anova

Berikut merupakan informasi pengelompokan dengan Metode Tukey dengan Confidence Level 95% yang ditunjukkan pada Gambar 6.15

Tukey Pairwise Comparisons				
Grouping Information Using the Tukey Method and 95% Confidence				
Factor	N	Mean	Grouping	
TIT 1 T1	102	6.892	A	
TIT 1 T0	102	6.382	A	
TIT 1 T3	102	6.284	A	
TIT 1 T2	102	5.990	A	

Means that do not share a letter are significantly different.

**Gambar 6.15** Informasi pengelompokan sampel pada *Topic Intrusion task 1*

Untuk kasus uji ANOVA *Topic Intrusion task 1*, seluruh topik digolongkan pada huruf A, yang artinya nilai rata-rata untuk setiap topik dinyatakan sama

Gambar 6.16 menunjukkan perbandingan berpasangan dari masing masing topik dalam uji ANOVA *Topic Intrusion task 1*

## Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
TIT 1 T1 - TIT 1 T0	0.510	0.408	(-0.537, 1.557)	1.25	0.595
TIT 1 T2 - TIT 1 T0	-0.392	0.408	(-1.439, 0.655)	-0.96	0.771
TIT 1 T3 - TIT 1 T0	-0.098	0.408	(-1.145, 0.949)	-0.24	0.995
TIT 1 T2 - TIT 1 T1	-0.902	0.408	(-1.949, 0.145)	-2.21	0.120
TIT 1 T3 - TIT 1 T1	-0.608	0.408	(-1.655, 0.439)	-1.49	0.443
TIT 1 T3 - TIT 1 T2	0.294	0.408	(-0.753, 1.341)	0.72	0.889

Individual confidence level = 98.94%

**Gambar 6.16** Perbandingan berpasangan setiap sampel untuk *Topic Intrusion task 1*

Berdasarkan Gambar 6.16, perbedaan varian ditunjukkan dengan melakukan perbandingan pada nilai *p-value* dan nilai alpha ( $\alpha=0.05$ ), jika nilai *p-value* lebih tinggi dari nilai alpha, maka rata-rata kedua topik dinyatakan berbeda signifikan. Dengan demikian, perbandingan berpasangan untuk setiap topik telah menunjukkan kesesuaian dengan yang ditunjukkan dalam tabel informasi pengelompokan

### 6.5.3.3. Uji ANOVA *Topic Intrusion task 2*

#### Pernyataan hipotesis uji ANOVA *Topic Intrusion task 2*

$H_0$ : semua *Means* untuk setiap sampel sama

$H_1$ : terdapat setidaknya satu *Means* yang berbeda

#### Uji Anova

Berikut merupakan informasi pengelompokan dengan Metode Games-Howell dengan Confidence Level 95% yang ditunjukkan pada Gambar 6.17

Games-Howell Pairwise Comparisons				
Grouping Information Using the Games-Howell Method and 95% Confidence				
Factor	N	Mean	Grouping	
TIT 2 T1	122	6.828	A	
TIT 2 T3	122	6.418	A	
TIT 2 T0	122	6.254	A	
TIT 2 T2	122	4.787	B	
Means that do not share a letter are significantly different.				

**Gambar 6.17** Informasi pengelompokan sampel pada *Topic Intrusion task 2*

Untuk kasus uji ANOVA *Topic Intrusion task 2*, yang tergolong dalam kelompok huruf A adalah topik 1 dan topik 3, sementara yang tergolong dalam kelompok huruf B adalah topik 0, 2

Gambar 6.18 menunjukkan perbandingan berpasangan dari masing masing topik dalam uji ANOVA *Word Intrusion Task*

Games-Howell Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
TIT 2 T1 - TIT 2 T0	0.574	0.324	(-0.265, 1.413)	1.77	0.290
TIT 2 T2 - TIT 2 T0	-1.467	0.272	(-2.171, -0.763)	-5.40	0.000
TIT 2 T3 - TIT 2 T0	0.164	0.315	(-0.646, 0.974)	0.52	0.954
TIT 2 T2 - TIT 2 T1	-2.041	0.293	(-2.799, -1.283)	-6.96	0.000
TIT 2 T3 - TIT 2 T1	-0.410	0.334	(-1.267, 0.447)	-1.23	0.609
TIT 2 T3 - TIT 2 T2	1.631	0.283	( 0.898, 2.365)	5.75	0.000

**Gambar 6.18** Perbandingan berpasangan setiap sampel untuk *Topic Intrusion task 2*

Berdasarkan Gambar 6.18, perbedaan varian ditunjukkan dengan melakukan perbandingan pada nilai *p-value* dan nilai alpha ( $\alpha=0.05$ ), jika nilai *p-value* lebih tinggi dari nilai alpha, maka rata-rata kedua topik dinyatakan berbeda signifikan. Dengan demikian, perbandingan berpasangan untuk setiap topik telah menunjukkan kesesuaian dengan yang ditunjukkan dalam tabel informasi pengelompokan

#### 6.5.3.4. Kesimpulan Uji ANOVA

Berdasarkan uji ANOVA yang telah dilakukan, analisis terhadap kualitas pemodelan topik hingga tingkat topik menghasilkan dua kesimpulan. Pertama, topik 1 merupakan topik yang paling mudah diinterpretasi oleh manusia, hal ini didukung dengan nilai rata-rata topik 1 yang paling tinggi baik untuk metode WIT, TIT 1 maupun TIT 2, sementara topik 2 merupakan topik yang paling sulit diinterpretasi oleh manusia, hal ini didukung dengan nilai rata-rata topik 2 yang paling rendah pada metode TIT 1 dan TIT 2, dan kedua terendah pada metode WIT. Kedua, dari keempat topik yang dihasilkan, dapat diketahui bahwa derajat pemahaman responden terhadap topik untuk metode WIT dan TIT 2 dapat dibagi ke dalam dua tingkat, sementara untuk metode TIT 1, hanya dalam satu tingkat. Dengan kata lain, topik-topik yang berada dalam kelompok A dapat dikatakan sebagai kelompok topik yang memiliki performa baik sebagai classifier, sementara topik-topik yang berada dalam kelompok B dapat dikatakan sebagai kelompok topik yang memiliki performa kurang baik

*Halaman ini sengaja dikosongkan*

## BAB VII

### KESIMPULAN DAN SARAN

Pada bab ini dibahas mengenai kesimpulan dari semua proses yang telah dilakukan dan saran yang dapat diberikan untuk pengembangan yang lebih baik.

#### 7.1. Kesimpulan

Kesimpulan yang didapatkan dari proses pengerjaan tugas akhir yang telah dilakukan antara lain:

1. Telah dilakukan berbagai eksperimen dalam proses *Topic modeling* dengan metode *LDA*. Eksperimen yang terkait jumlah *passes* menghasilkan 10 *passes* sebagai jumlah *passes* terbaik berdasarkan nilai *perplexity* yang telah stabil pada iterasi ke 10 untuk setiap distribusi topik yang terpilih secara acak.
2. Eksperimen yang terkait dengan jumlah topik dalam model dibagi menjadi dua eksperimen, yaitu *Topic modeling* dengan *Stem* dan *Topic modeling* tanpa *Stem*. Berdasarkan hasil eksperimen ini, dihasilkan beberapa kesimpulan. Kesimpulan pertama, bahwa 4 topik merupakan jumlah topik terbaik dalam membentuk topik model berdasarkan analisis nilai *perplexity* dan dapat dikatakan cukup stabil berdasarkan nilai standar deviasinya. Kesimpulan kedua, proses *Stemming* dalam kasus ini memberikan pengaruh positif terhadap pembentukan model yang ditunjukkan dengan penurunan nilai *perplexity* sebesar 64.51.
3. Eksperimen terkait klasifikasi dokumen ke dalam topik telah dilakukan baik untuk model dengan *Stem*, maupun model tanpa *Stem*. Hasil eksperimen klasifikasi dokumen menunjukkan hasil yang meyakinkan yang ditunjukkan dengan distribusi probabilitas dokumen terhadap topik untuk model dengan atau tanpa *Stem* secara mayoritas berada pada rentang 0.5 hingga 0.99.

4. Dalam mengukur tingkat koherensi topik yang menjadi luaran *Topic modeling*, metode yang digunakan adalah *Word Intrusion task* dan *Topic Intrusion Task*. Tahap-tahap yang dilakukan adalah merancang sistematisasi dan materi kuesioner, kemudian melakukan analisis hasil tanggapan responden dengan uji hipotesis dan uji ANOVA
5. Berdasarkan uji hipotesis yang dilakukan melalui uji *variance* dan uji *Means*, terdapat dua kesimpulan yang dapat diambil. Pertama, metode *Word Intrusion task* secara inheren mampu diinterpretasi lebih baik oleh responden. Kedua, nilai rata-rata metode WIT, TIT 1, dan TIT 2 dengan *Stem* memberikan hasil yang lebih baik. Hal ini menunjukkan bahwa proses *Stemming* berpengaruh positif terhadap pemodelan yang dilakukan dalam kasus ini, namun berdasarkan uji *variance* dan uji *Means*, perbedaan ini tidak signifikan.
6. Berdasarkan uji ANOVA yang telah dilakukan, analisis terhadap kualitas pemodelan topik hingga tingkat topik menghasilkan dua kesimpulan. Pertama, topik 1 merupakan topik yang paling mudah diinterpretasi oleh manusia, sementara topik 2 merupakan topik yang paling sulit diinterpretasi oleh manusia. Kedua, dari keempat topik yang dihasilkan, dapat diketahui bahwa derajat pemahaman responden terhadap topik untuk metode WIT dan TIT 2 dapat dibagi ke dalam dua tingkat, sementara untuk metode TIT 1, hanya dalam satu tingkat.

## 7.2. Saran

Dari pengerjaan tugas akhir ini, terdapat beberapa saran untuk pengembangan penelitian ke depan.

1. Data yang digunakan dalam penelitian ini bersumber dari status *Facebook fanpage* dan *twitter* yang menggunakan kebahasaan berita media sosial. Pada pengembangannya, dapat diujicobakan dataset dari



akun media sosial lain yang terpercaya ataupun dataset yang diambil dari komentar pada status *Facebook fanpage* yang cenderung lebih banyak menggunakan karakteristik bahasa informal sehari-hari.

2. Berdasarkan pengamatan yang dilakukan penulis, diperoleh dugaan sementara bahwa terdapat perbedaan karakteristik data yang bersumber dari facebook dan twitter, baik dari panjang karakter per dokumen dan sifat kebahasaan yang diterapkan, sehingga hal ini dapat menjadi poin untuk dilakukan pengembangan ke depan
3. Dari hasil pemodelan topik, terdapat kata-kata yang tergolong sebagai nama lokasi dalam topik, sehingga untuk memperoleh hasil yang lebih optimal, diperlukan suatu pendeteksian nama lokasi sebagai penyaringan data sebelum dilakukan pemodelan, demikian pula untuk merek dagang. Terdapat pula kata-kata yang berupa singkatan yang memiliki arti sama namun dalam pemodelan topik menjadi dua entitas yang berbeda, seperti *jln* dengan *jalan* dan *yg* dengan *yang*. Untuk itu diperlukan suatu normalisasi sebelum dilakukan pemodelan topik

*Halaman ini sengaja dikosongkan*

## DAFTAR PUSTAKA

- [1] P. K. Surabaya, "Situs Web Resmi Kota Surabaya," Pemerintah Kota Surabaya, [Online]. Available: <http://sparkling.surabaya.go.id/>. [Accessed 19 Oktober 2016].
- [2] D. I. P. Hadi, "Penyiaran Interaktif dan Kepentingan Publik," 2014.
- [3] R. S. Surabaya, "Suara Surabaya Official Website," [Online]. Available: <http://www.suarasurabaya.net/>. [Accessed 19 Oktober 2016].
- [4] D. M. Blei, "Latent Dirichlet Allocation," *Machine Learning Research* 3, pp. 933-1022, 2003.
- [5] J. H. L. K. G. T. B. David Newman, "Automatic Evaluation of Topic Coherence," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*, pp. 100-108, 2010.
- [6] M. H. Basri, "IDENTIFIKASI TOPIK INFORMASI PUBLIK MEDIA SOSIAL DI KOTA SURABAYA BERDASARKAN KLASTERISASI TEKS PADA TWITTER DENGAN MENGGUNAKAN ALGORITMA K-MEANS," 2016.
- [7] H. Z. Zhou Tong, "A Text Mining Research Based On LDA Topic Modelling," *Jodrey School of Computer Science, Acadia University, Wolfville, NS, Canada*, vol. 10.5121/csit.2016.60616, p. 201–210, 2016.
- [8] A. Gaur, "Topic Models As A Novel Approach To Identify Themes In Content Analysis: The Example Of Organizational Research Methods," 2015.
- [9] N. L. D. F. d. D. S. Willyh Hariardi, "Automatic Summarization from Indonesian Hashtag on Twitter Using TF-IDF and Phrase Reinforcement Algorithm," *International Workshop on Computer*

*Science and Engineering (WCSE 2016)*, pp. 575-579, 2016.

- [10] N. C. T. B. Jey Han Lau, "On-line Trend Analysis with Topic Models: #twitter trends detection topic model online," *Proceedings of COLING 2012: Technical Papers*, p. 1519–1534, 2012.
- [11] P. S. Radim Rehurek, "Radim Rehurek dan Petr Sojka," *Valletta, Malta, Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, pp. 46--50, 2010.
- [12] D. M. Blei, "Probabilistic Topic Model," *communications of the acm*, vol. 55, 2012.
- [13] M. R. Brett, "Journal of Digital Humanities," Desember 2012. [Online]. Available: <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>.
- [14] J. C. Campbell, A. Hindle and a. E. Stroulia, "Latent Dirichlet Allocation: Extracting Topics," 2014.
- [15] J.-F. Yeh, Y.-S. Tan and C.-H. Lee, "Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet Allocation," *Neurocomputing*, 2016.
- [16] P. K. D. A. D. B. Keith Stevens, "Exploring Topic Coherence over many models and many topics," *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, p. 952–961, 2012.
- [17] J. H. L. K. G. a. T. B. David Newman, "Automatic Evaluation of Topic Coherence," *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, p. 100–108, 2010.
- [18] J. B.-G. C. W. S. G. D. M. B. Jonathan Chang, "Reading Tea Leaves: How Humans Interpret Topic Models,"

- in *Neural Information Processing Systems*, Vancouver, BC, 2009.
- [19] H. Tanhana Jukangko, "Peran Radio Suara Surabaya Sebagai Media Pendidikan untuk Tertib Berlalu Lintas bagi Warga Kota Surabaya," *Kajian Moral dan Kewarganegaraan*, vol. Vol 3 Nomer 1, 2013.
- [20] R. S. Surabaya, "Official Facebook Page Radio Suara Surabaya," [Online]. Available: <https://www.facebook.com/e100ss/about/>. [Accessed 3 November 2016].
- [21] I. P. Hadi, "Khalayak Maya Dalam Media Online," *Jurnal Ilmiah SCRIPTURA*, vol. 1, 2007.
- [22] F. L. Desi Yoanita, "Akurasi dalam Jurnalisma Warga pada Radio Suara Surabaya`," *Jurnal SCRIPTURA*, vol. 4, pp. 47-53, 2014.
- [23] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands*, 2003.
- [24] cslu.ogi.edu, "<http://www.cslu.ogi.edu/>," [Online]. Available: <http://www.cslu.ogi.edu/>. [Accessed 17 Desember 2016].
- [25] M. Muslich, *Garis-garis Besar Tata Bahasa Baku Bahasa Indonesia*, Bandung: Refika Aditama, 2010.
- [26] A. Chaer, *Tata Bahasa Praktis Bahasa Indonesia*, Jakarta: Barata Karya Aksara, 2008.

*Halaman ini sengaja dikosongkan*

## BIODATA PENULIS



Penulis lahir di Denpasar, tanggal 10 April 1995 dengan nama I Made Kusnanta Bramantya Putra sebagai anak kedua dari pasangan Putu Arya Mertasana dan Made Mahayuni. Penulis telah menempuh beberapa pendidikan formal yaitu; SD Saraswati 4 Denpasar, SMP Negeri 7 Denpasar, dan SMA Negeri 1 Denpasar

Pada tahun 2013 pasca kelulusan SMA, penulis melanjutkan pendidikan dengan jalur SNMPTN Undangan di Departemen Sistem Informasi Fakultas Teknologi Informasi dan Komunikasi – Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5213100035.

Selama menjadi mahasiswa, penulis aktif mengikuti berbagai kegiatan kemahasiswaan seperti beberapa kepanitiaan serta organisasi mahasiswa, ditunjukkan dengan amanah sebagai staf Pengembangan Sumber Daya Mahasiswa (PSDM) di Himpunan Mahasiswa Sistem Informasi (HMSI) dan Tim Pembina Kerohanian Hindu ITS (TPKH-ITS) pada tahun kedua, dan meneruskan sebagai Wakil Ketua HMSI pada tahun ketiga. Selain itu, penulis juga aktif di Latihan Keterampilan Manajemen Mahasiswa (LKMM) hingga LKMM Tingkat Menengah baik sebagai peserta dan pemandu. Di bidang akademik, penulis aktif menjadi asisten praktikum pada beberapa mata kuliah seperti Bahasa Pemrograman, dan Perencanaan Sumber Daya Perusahaan.

Pada tahun keempat, karena penulis memiliki ketertarikan di bidang data *mining*, maka penulis mengambil bidang minat Akuisisi Data dan Diseminasi Informasi (ADDI). Penulis dapat dihubungi melalui *email* di kusnantamade@gmail.com.

*Halaman ini sengaja dikosongkan*



## LAMPIRAN A

### A.1. *Stopwords*

<i>Stopwords</i>	<i>Jenis Kata</i>	<i>Stopwords</i>	<i>Jenis Kata</i>
ada	konjungsi	adalah	konjungsi
adanya	konjungsi	adapun	partikel, konjungsi
aduh	interjeksi	aduhai	interjeksi
agak	adverbia	agaknya	adverbia
agar	partikel, konjungsi	agar supaya	konjungsi
ah	interjeksi	ai	interjeksi
akan	partikel	akan hal	konjungsi
akan tetapi	konjungsi	akankah	partikel
akhir	verba	akhiri	verba
akhirnya	verba	aku	pronomia, kata ganti orang
akulah	pronomia	alhamdulillah	interjeksi
amat	adverbia	amatlah	adverbia
amboi	interjeksi	anda	verba
andai	konjungsi	andaikata	konjungsi
andalah	verba	antar	partikel, preposisi
antara	preposisi	antaranya	partikel
anu	kata ganti penunjuk	apa	pronomia, kata ganti penunjuk
apaan	pronomia	apabila	partikel, konjungsi, kata ganti penunjuk
apakah	pronomia	apalagi	pronomia
apatah	pronomia	artinya	konjungsi
asal	partikel, konjungsi	asalkan	partikel, konjungsi

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
astaga	interjeksi	astagfirullah	interjeksi
asyik	interjeksi	atas	preposisi
atau	partikel, konjungsi	ataukah	partikel
ataupun	partikel	awal	preposisi
awalnya	preposisi	ayo	interjeksi
bagai	partikel	bagaikan	partikel
bagaimana	pronomia, kata ganti penunjuk	bagaimanakah	pronomia
bagaimanapun	pronomia	bagi	partikel, preposisi
bagian	adverbia	bah	interjeksi, adverbia
bahkan	adverbia	bahwa	partikel, konjungsi
bahwasanya	partikel, konjungsi	baik	adjektiva
bakal	adverbia	bakalan	verba
balik	preposisi	bangsat	interjeksi
banyak	adjektiva	bapak	kata ganti orang
barang siapa	kata ganti penunjuk	baru	adjektiva
bawah	adverbia	beberapa	numeralia
bedebah	interjeksi	begini	pronomia, kata ganti penunjuk
beginian	adjektiva	beginikah	pronomia
beginilah	pronomia	begitu	adverbia, kata ganti penunjuk
begitukah	adverbia	begitulah	adverbia
begitupun	adverbia	bekerja	verba
belakang	adverbia	belakangan	adverbia
beliau	kata ganti orang	belum	adverbia

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
belumlah	adverbia	benar	adjektiva
benarkah	adjektiva	benarlah	adjektiva
berada	verba	berakhir	verba
berakhirlah	verba	berakhirnya	konjungsi
berapa	pronomia, kata ganti penunjuk	berapakah	pronomia
berapalah	pronomia	berapapun	pronomia
berarti	verba	berawal	verba
berbagai	verba	berdatangan	verba
berhubung	konjungsi	beri	verba
berikan	verba	berikut	adjektiva
berikutnya	adjektiva	berjumlah	verba
berkali-kali	adverbia	berkat	preposisi
berkata	verba	berkehendak	verba
berkeinginan	verba	berkenaan	verba
berlainan	verba	berlalu	verba
berlangsung	verba	berlebihan	adjektiva
bermacam	adjektiva	bermacam-macam	adjektiva
bermaksud	verba	bermula	verba
bersama	verba, preposisi	bersama sama	preposisi
bersama-sama	verba	bersiap	verba
bersiap-siap	verba	bertanya	verba
bertanya-tanya	verba	berturut	adverbia
berturut-turut	adverbia	bertutur	verba
berujar	verba	berupa	verba
besar	adjektiva	beserta	preposisi
betul	adjektiva	betulkah	adjektiva
biar	konjungsi, konjungsi	biarpun	konjungsi

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
biarpun begitu	konjungsi	biarpun demikian	konjungsi
biasa	adjektiva	biasanya	adjektiva
bila	partikel, konjungsi	bilakah	partikel
bilamana	kata ganti penunjuk	bisa	verba
bisakah	verba	boleh	partikel
bolehhkah	partikel	bolehlah	partikel
brengsek	interjeksi	buat	partikel, preposisi
bukan	adverbia	bukankah	pronomia
bukanlah	adverbia	bukannya	adverbia
bulan	adverbia	bung	kata ganti orang
cara	adjektiva	caranya	adjektiva
cih	interjeksi	cis	interjeksi
cukup	adjektiva	cukupkah	adjektiva
cukuplah	adjektiva	cuma	adverbia
dahulu	adverbia	daku	kata ganti orang
dalam	partikel	dalam pada itu	konjungsi
dan	partikel, konjungsi	dang	artikel
dapat	adverbia	dari	preposisi, partikel
dari mana	kata ganti penunjuk	daripada	partikel, preposisi
datang	verba	dekat	adjektiva, preposisi
demi	partikel, konjungsi	demikian	pronomia, konjungsi
demikianlah	pronomia	dengan	preposisi, partikel, konjungsi
dengan begitu	konjungsi	dengan demikian	konjungsi
depan	adverbia	di	preposisi, partikel

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
di mana	kata ganti penunjuk	dia	pronomia, kata ganti orang
diakhiri	verba	diakhirinya	verba
dialah	pronomia	diantara	verba
diantaranya	verba	diberi	verba
diberikan	verba	diberikannya	verba
dibuat	verba	dibuatnya	verba
didapat	verba	didatangkan	verba
digunakan	verba	diibaratkan	verba
diibaratkannya	verba	diingat	verba
diingatkan	verba	diinginkan	verba
dijawab	verba	dijelaskan	verba
dijelaskannya	verba	dikarenakan	verba
dikatakan	verba	dikatakannya	verba
dikau	kata ganti orang	dikerjakan	verba
diketahui	verba	diketahuinya	verba
dikira	verba	dilakukan	verba
dilalui	verba	dilihat	verba
dimaksud	verba	dimaksudkan	verba
dimaksudkannya	verba	dimaksudnya	verba
diminta	verba	dimintai	verba
dimisalkan	verba	dimulai	verba
dimulailah	verba	dimulainya	verba
dimungkinkan	verba	dini	adjektiva
dipastikan	verba	diperbuat	verba
diperbuatnya	verba	dipergunakan	verba
diperkirakan	verba	diperlihatkan	verba
diperlukan	verba	diperlukannya	verba

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
dipersoalkan	verba	dipertanyakan	verba
dipunyai	verba	diri	verba
dirinya	kata ganti orang	disampaikan	verba
disebut	verba	disebutkan	verba
disebutkannya	verba	disini	adverbia
disinilah	adverbia	ditambahkan	verba
ditandaskan	verba	ditanya	verba
ditanyai	verba	ditanyakan	verba
ditegaskan	verba	ditujukan	verba
ditunjuk	verba	ditunjuki	verba
ditunjukkan	verba	ditunjukkannya	verba
ditunjuknya	verba	dituturkan	verba
dituturkannya	verba	diucapkan	verba
diucapkannya	verba	diungkapkan	verba
dong	partikel	dua	numeralia
dulu	adverbia	e100	reserved word
e100ss	reserved word	eh	interjeksi
empat	numeralia	enggak	adverbia
enggaknya	adverbia	engkau	kata ganti orang
entah	adverbia, konjungsi	entahlah	adverbia
guna	preposisi	gunakan	verba
haa	interjeksi	hadapan	preposisi
hai	interjeksi	hal	kata ganti
hampir	adverbia	hang	artikel
hanya	adverbia	hanyalah	adverbia
hari	adverbia	harus	adverbia
haruslah	adverbia	harusnya	adverbia

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
he	interjeksi	hem	interjeksi
hendak	partikel	hendaklah	adverbia
hendaknya	partikel	hingga	partikel, konjungsi, preposisi
hm	reserved word	ia	pronomia, kata ganti orang
ialah	partikel	ibarat	partikel
ibaratkan	verba	ibaratnya	partikel
ibu	kata ganti orang	ih	interjeksi
ikut	verba	ingat	verba
ingat-ingat	verba	ingin	partikel
inginkah	verba	inginkan	verba
ini	pronomia, kata ganti penunjuk	inikah	pronomia
inilah	pronomia	insya Allah	interjeksi
isih	interjeksi	itu	pronomia, kata ganti penunjuk
itukah	pronomia	itulah	pronomia
jadi	verba	jadilah	verba
jadinya	konjungsi	jangan	partikel
janganakan	partikel, konjungsi	janganlah	partikel
jauh	adjektiva	jawab	verba
jawaban	verba	jawabnya	verba
jelas	adjektiva	jelaskan	verba
jelastah	adjektiva	jelastanya	verba
jika	partikel, konjungsi	jikalau	partikel, konjungsi
juga	adverbia	jumlah	numeralia

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
jumlahnya	numeralia	justru	adverbia
kala	adverbia	kalau	partikel, konjungsi
kalaupun	partikel	kalaupun	partikel
kalian	pronomia, kata ganti orang	kami	pronomia, kata ganti orang
kamilah	pronomia	kamu	pronomia
kamu sekalian	kata ganti orang	kamulah	pronomia
kan	partikel	kapan	partikel, kata ganti penunjuk
kapankah	partikel	kapanpun	partikel
karena	preposisi, partikel, konjungsi	karenanya	partikel
kasus	verba	kata	verba
katakan	verba	katakanlah	verba
katanya	kata kerja, kata ganti orang	kau	kata ganti orang
ke	preposisi, partikel	ke mana	kata ganti penunjuk
keadaan	adverbia	kebetulan	adverbia
kecil	adjektiva	kecuali	preposisi
kecuali itu	konjungsi	kedua	numeralia
keduanya	numeralia	keinginan	verba
kelamaan	adjektiva	kelihatan	verba
kelihatannya	verba	kelima	numeralia
keluar	verba	kembali	verba
kemudian	partikel	kemungkinan	verba
kemungkinannya	verba	kenapa	pronomia
kendati	konjungsi	kepada	partikel, preposisi



<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
kepadanya	partikel	keparat	interjeksi
kesampaian	verba	keseluruhan	verba
keseluruhannya	verba	keterlaluhan	adjektiva
ketika	konjungsi	khususnya	adverbia
kini	adverbia	kinilah	adverbia
kira	verba	kira-kira	adverbia
kiranya	adverbia	kita	pronomia, kata ganti orang
kitalah	pronomia	kok	partikel
ku	kata ganti orang	kurang	adverbia
lagi	adverbia, konjungsi	lagi pula	konjungsi
lagian	adverbia	lah	partikel
lain	adjektiva	lainnya	adjektiva
lalu	verba	lama	adjektiva
lamanya	verba	lanjut	adjektiva
lanjutnya	verba	lantaran	konjungsi
lantas	konjungsi	lebih	adjektiva
lewat	partikel, preposisi	lho	interjeksi
lima	numeralia	lo	interjeksi
luar	preposisi	macam	adverbia
maka	partikel, konjungsi	makanya	partikel, konjungsi
makin	adverbia	malah	adverbia
malahan	adverbia, konjungsi	mampu	adjektiva
mampukah	adjektiva	mana	pronoun, kata ganti penunjuk
manakala	partikel, konjungsi	manalagi	partikel

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
masa	adverbial	masalah	verba
masalahnya	verba	masih	adverbial
masihkah	adverbial	masing	pronomial
masing-masing	pronomial, kata ganti penunjuk	masyaallah	interjeksi
mau	adverbial	maupun	partikel, konjungsi
melainkan	verba	melakukan	verba
melalui	verba, preposisi	melihat	verba
melihatnya	verba	memang	adverbial
memastikan	verba	memberi	verba
memberikan	verba	membuat	verba
memerlukan	verba	memihak	verba
meminta	verba	memintakan	verba
memisalkan	verba	memperbuat	verba
mempergunakan	verba	memperkirakan	verba
memperlihatkan	verba	mempersiapkan	verba
mempersoalkan	verba	mempertanyakan	verba
mempunyai	verba	memulai	verba
memungkinkan	verba	menaiki	verba
menambahkan	verba	menandaskan	verba
menanti	verba	menantikan	verba
menanti-nanti	verba	menanya	verba
menanyai	verba	menanyakan	verba
mendapat	verba	mendapatkan	verba
mendatang	adjektiva	mendatangi	verba
mendatangkan	verba	menegaskan	verba
mengakhiri	verba	mengapa	pronomial, kata ganti penunjuk

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
mengatakan	verba	mengatakannya	verba
mengenai	partikel, konjungsi	mengerjakan	verba
mengetahui	verba	menggunakan	verba
menghendaki	verba	mengibaratkan	verba
mengibaratkannya	verba	mengingat	verba, preposisi
mengingatkan	verba	menginginkan	verba
mengira	verba	mengucapkan	verba
mengucapkannya	verba	mengungkapkan	verba
menjadi	verba	menjawab	verba
menjelang	preposisi	menjelaskan	verba
menuju	verba, preposisi	menunjuk	verba
menunjuki	verba	menunjukkan	verba
menunjuknya	verba	menurut	partikel, preposisi
menuturkan	verba	menyampaikan	verba
menyangkut	verba	menyatakan	verba
menyebutkan	verba	menyeluruh	verba
menyiapkan	verba	merasa	verba
mereka	pronomia, kata ganti orang	merekalah	pronomia
merupakan	verba	meski	partikel, konjungsi
meskipun	partikel, konjungsi	meyakini	verba
meyakinkan	verba	minta	verba
mirip	adverbia	misal	verba
misalkan	verba	misalnya	verba
mu	kata ganti orang	mula	verba
mulai	verba	mulailah	verba

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
mulanya	verba	mungkin	adverbia
mungkin	adverbia	nah	partikel
naik	verba	namun	partikel, konjungsi
nanti	adverbia	nantinya	adverbia
nya	kata ganti orang	nyaris	adverbia
nyatanya	adjektiva	odp	reserved word
oh	interjeksi	oleh	preposisi, partikel
oleh karena	preposisi, konjungsi	oleh karena itu	konjungsi
oleh sebab	preposisi	olehnya	partikel
olehpada	preposisi	pada	preposisi, partikel
padahal	partikel	padanya	partikel
pak	pronomia	paling	adverbia
panjang	adjektiva	pantas	adjektiva
para	partikel, artikel	pasti	adjektiva
pastilah	adjektiva	penting	adjektiva
pentingnya	adjektiva	per	particle
percuma	adverbia	perlu	adverbia
perlukah	adverbia	perlunya	verba
pernah	adverbia	persoalan	verba
pertama	numeralia	pertama-tama	adverbia
pertanyaan	verba	pertanyakan	verba
pihak	kata ganti	pihaknya	kata ganti
pr	reserved word	pukul	verba
pula	partikel, konjungsi	pun	partikel
punya	verba	rasa	verba

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
rasanya	verba	rata	adverbia
rt	reserved word	rupanya	verba
saat	adverbia	saatnya	adverbia
saja	adverbia	sajalah	adverbia
saling	adverbia	sama	adjektiva, preposisi
sama-sama	adjektiva	sambil	partikel, konjungsi
sampai	verba, konjungsi	sampai dengan	preposisi
sampai ke	preposisi	sampaikan	verba
sampai-sampai	verba, konjungsi	sana	kata ganti penunjuk
sang	artikel	sangat	adverbia
sangatlah	adverbia	satu	numeralia
saya	pronomia, kata ganti orang	sayalah	pronomia
se	partikel	seakan akan	konjungsi
seandainya	konjungsi	sebab	preposisi, partikel, konjungsi
sebabnya	partikel	sebagai	partikel
sebagaimana	partikel, konjungsi	sebagaimanalaksana	konjungsi
sebagainya	partikel	sebagian	adverbia
sebaik	adjektiva	sebaik-baiknya	adverbia
sebaiknya	adverbia	sebaliknya	adverbia, konjungsi
sebanyak	numeralia	sebegini	numeralia
sebegitu	numeralia	sebelum itu	konjungsi
sebelumnya	adverbia	sebenarnya	adverbia
seberapa	numeralia	sebesar	adjektiva
sebetulnya	adverbia	sebisanya	adverbia

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
sebuah	numeralia	sebut	verba
sebutlah	verba	sebutnya	verba
secara	partikel	secukupnya	adjektiva
sedang	partikel	sedangkan	partikel
sedari	preposisi	sedemikian	pronomia
sedikit	adjektiva	sedikitnya	adverbia
seenaknya	adjektiva	segala	adjektiva
segalanya	adjektiva	segera	adverbia
sehabis	konjungsi	seharusnya	adverbia
sehingga	partikel, konjungsi	seingat	adverbia
sejak	preposisi, partikel, konjungsi	sejauh	adverbia
sejenak	adverbia	sejumlah	adverbial, numeralia
sekadar	adverbia	sekadarnya	adverbia
sekali	adverbia	sekalian	numeralia
sekaligus	adverbia	sekali-kali	adverbia
sekalipun	partikel, konjungsi	sekarang	adverbia
sekecil	adjektiva	sekeliling	preposisi
seketika	adverbia	sekiranya	verba, konjungsi
sekitar	adverbia, preposisi	sekitarnya	adverbia, preposisi
sekurang- kurangnya	adverbia	sekurangnya	adverbia
sela	adverbia	selagi	konjungsi
selain	partikel, preposisi	selain dari	preposisi
selain itu	konjungsi	selaku	partikel
selalu	adverbia	selama	preposisi, konjungsi
selama-lamanya	adjektiva	selamanya	adjektiva

<b><i>Stopwords</i></b>	<b><i>Jenis Kata</i></b>	<b><i>Stopwords</i></b>	<b><i>Jenis Kata</i></b>
selanjutnya	adverbia, konjungsi	selesai	konjungsi
seluruh	numeral	seluruhnya	numeral
semacam	adverbia, preposisi	semakin	adverbia
semampu	adjektiva	semampunya	adjektiva
semasa	adverbia	semasih	adverbia
semata	adverbia	semata-mata	adverbia
semaunya	adverbia	sembari	konjungsi
sementara	partikel, konjungsi	semisal	adverbia
semisalnya	adverbia	sempat	adverbia
semua	numeralia	semuanya	adverbia
semula	adverbia	sendiri	adverbia, kata ganti penunjuk
sendirian	pronomia	sendirinya	adverbia
seolah	verba	seolah olah	konjungsi
seolah-olah	adverbia	seorang	numeralia
sepanjang	numeralia, preposisi	sepantasnya	adjektiva
sepantasnyalah	adjektiva	seperlunya	adverbia
seperti	partikel, konjungsi	sepertinya	partikel
sepihak	adverbia	seputar	preposisi
serasa	konjungsi	seraya	konjungsi, konjungsi
sering	adverbia	seringnya	adverbia
serta	partikel	serupa	verba
sesaat	preposisi	sesama	adverbia
sesampai	partikel	sesegea	adverbia
sese kali	adverbia	seseorang	kata ganti penunjuk

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
sesuatu	pronomia, kata ganti penunjuk	sesuatunya	pronomia
sesudah	partikel, konjungsi	sesudah itu	konjungsi
sesudahnya	partikel	sesungguhnya	konjungsi
setelah	adverbia, konjungsi	setelah itu	konjungsi
setempat	adverbia	setengah	numeralia
seterusnya	adverbia	setiap	numeralia
setiba	partikel	setibanya	adverbia
setidaknya	adverbia	setidak-tidaknya	adverbia
setinggi	adjektiva	seusai	partikel
sewaktu	konjungsi	si	artikel
sialan	interjeksi	siap	verba
siapa	pronomia, kata ganti penunjuk	siapakah	pronomia
siapapun	pronomia	sini	adverbia, kata ganti penunjuk
sinilah	adverbia	situ	kata ganti penunjuk
soal	adverbia	soalnya	adverbia
sri	artikel	ss	reserved word
ssinfo	reserved word	suatu	pronomia
sudah	adverbia	sudahkah	adverbia
sudahlah	adverbia	sungguh	konjungsi
supaya	partikel, konjungsi	syukur	interjeksi
tadi	adverbia	tadinya	adverbia
tahu	verba	tahun	adverbia
tak	adverbia	tambah	verba



<i><b>Stopwords</b></i>	<i><b>Jenis Kata</b></i>	<i><b>Stopwords</b></i>	<i><b>Jenis Kata</b></i>
tambahnya	verba	tampak	verba
tampaknya	verba	tandas	adjectiva
tandasnya	verba	tanpa	adverbia, preposisi
tanya	verba	tanyakan	verba
tanyanya	verba	tapi	partikel
tatkala	konjungsi	tatkalatengah	konjungsi
tegas	verba	tegasnya	verba
telah	adverbia	tempat	preposisi
tengah	adverbia, preposisi	tentang	preposisi, partikel
tentu	adjektiva	tentulah	adjektiva
tentunya	adverbia	tepat	adjektiva
terakhir	adjektiva	terasa	verba
terbanyak	adjektiva	terdahulu	adverbia
terdapat	verba	terdiri	verba
terhadap	partikel, preposisi	terhadapnya	partikel
teringat	verba	teringat-ingat	verba
terjadi	verba	terjadilah	verba
terjadinya	adverbia	terkira	verba
terlalu	adverbia	terlebih	adverbia
terlihat	verba	termasuk	verba
ternyata	verba	tersampaikan	verba
tersebut	verba	tersebutlah	verba
tertentu	adjektiva	tertuju	verba
terus	adverbia	terutama	adverbia
tetap	adjektiva	tetapi	partikel, konjungsi
tiap	adjektiva	tiba	verba
tiba-tiba	adverbia	tidak	adverbia

<b><i>Stopwords</i></b>	<b>Jenis Kata</b>	<b><i>Stopwords</i></b>	<b>Jenis Kata</b>
tidak hanya	konjungsi	tidakkah	adverbia
tidaklah	adverbia	tiga	numeralia
tinggi	adjektiva	tinimbang	preposisi
toh	partikel	tunjuk	verba
turut	verba	tutur	verba
tuturnya	verba	ucap	verba
ucapnya	verba	ujar	verba
ujarnya	verba	umpamanya	konjungsi
umum	adjektiva	umumnya	adverbia
ungkap	verba	ungkapnya	verba
untuk	partikel, preposisi, konjungsi	usah	verba
usai	verba	waduh	partikel
wah	partikel	wahai	partikel, interjeksi
wahai	interjeksi	waktu	konjungsi
waktunya	konjungsi	walau	partikel, konjungsi
walaupun	partikel, konjungsi	wong	pronomia
ya	interjeksi	ya ampun	interjeksi
yaitu	partikel	yakin	adjektiva
yakni	partikel	yang	partikel, konjungsi
yang mana	kata ganti penunjuk	yg	partikel

## A.2. Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik dengan *Stemming*

**Tabel 0.1 Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik dengan *Stemming***

Iterasi	Jumlah Topik								
	2	3	4	5	6	7	8	9	10
1	233.6	219.1	220.5	222.4	236.8	242.3	250	247.2	252.8
2	232.2	220	213.8	224.6	233.6	242.1	251.8	246.3	247.3
3	232.2	218.7	212.1	213.3	238.2	246.1	248.2	248.2	248.7
4	233.3	219.1	207.6	213.1	242.7	242.9	244.7	252.1	257.1
5	232.4	217.7	215.3	220.4	238.1	242.1	242.7	249.9	250.6
6	232.2	218.5	214.1	219.1	234.7	247.5	245.8	255	247.8
7	234.6	217.2	211.5	225.5	233.7	244.7	245.4	256.6	248.9
8	232.4	218.8	213.8	224.1	235	244.9	247.3	249.4	250.2
9	232.2	217.7	217.3	222	237.7	243.9	249	251	250.7
10	232.2	217.5	213.1	218.2	230.6	242.8	250.8	250	256.6
11	233.3	218.6	216.2	219.7	233.1	241.1	251.6	246.8	249.1
12	232.2	220	211.3	217.5	235.8	248.4	252.1	252.5	252.2
13	232.2	217.8	215.3	223.7	234.4	244.6	246	253.3	250.5
14	233.3	219.1	213.1	215.5	233.6	235.9	245.2	247	254.7
15	232.3	221.1	215.1	224.4	228.5	244.8	245.4	251.4	246.1
16	233.4	220.7	210.6	217.4	239.3	241.5	251.1	250.7	256.3
17	232.2	217	213.4	220.3	235	240.4	243.9	253.6	248.2
18	233.6	220	210.2	221.8	237.2	245.1	249.2	250.5	248.4
19	232.2	219.7	210.5	218.1	237	240.5	248.7	246.1	254
20	233.3	220.2	211.1	214.2	238	246.3	252.2	251.8	250.4
21	232.2	218.3	210.6	228.2	236.4	245	245.6	250.1	254.8
22	233.5	217	211.5	219.5	230.4	245.7	248.7	247	246.9

23	233.3	218.6	216.2	224	237.6	240.2	246.5	253.3	251.9
24	232.2	217.1	215	216.2	239.6	252.3	247.4	251.6	247.8
25	232.2	218.9	213.4	213.1	235.2	246.8	249.1	246.8	253.5
26	232.4	218.2	211.1	214	235.7	242.6	248.8	247.6	252
27	232.2	218	214.4	216.2	235.4	239.7	247	247.3	249.7
28	233.5	218.2	215.1	221.2	236.3	244.5	243.9	248.7	252.4
29	232.2	217.4	214.6	223.2	234.7	240.7	248.5	250.8	247.7
30	232.2	220.9	214.4	217.1	237.8	247.5	248	246.2	250.5

**Tabel 0.2 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik dengan Stemming (lanjutan)**

Iterasi	Jumlah Topik								
	14	18	22	26	30	34	38	42	46
1									
2	259.5	259.4	268.6	272	280.2	289.2	291	292.1	302.4
3	257.2	260.1	270.7	276.6	271.3	282.2	288.1	290.8	305.5
4	258.5	265.1	262.1	274.7	285	285.9	291.9	298.7	300.6
5	255.1	263	266.2	266.3	278.6	284.3	292.7	300.2	300.9
6	257.7	257.7	266.1	276.8	280	285.9	288	296.6	297.3
7	257.5	258.2	271	268.7	282	283.9	290.3	296.5	302.4
8	259.3	263.8	269.2	273.2	278.7	285.4	293.4	297	306.8
9	257.1	265	270.1	274.3	277.9	286.7	288.4	295.1	300.8
10	256.1	260.4	267.2	274.9	277	286.1	295.6	296.5	303
11	253.7	265.7	268.6	271.9	279.3	282.3	294.7	298.5	304.9
12	252.5	263.1	266.6	270.9	276	282.5	294.9	303.5	301
13	252	260.4	266.2	270.4	279.2	287.9	292.5	292.5	296.5
14	260.1	259.8	267.4	270.6	279.9	280.1	292.4	295.3	301.9
15	253.5	260.8	267.5	274.1	281.5	291.1	294.8	295.1	300.8
16	255.3	264.2	270.2	271.9	275.9	287	290.2	294.8	301.5

17	260	263.7	269.4	272.7	284.3	286	292.7	292.9	309.9
18	257.1	260.3	263.7	271.3	284.9	287.9	288.7	294.3	304.7
19	257.1	263	268.6	272.9	272.5	287.2	296.3	295.2	300.4
20	259	258.1	271.6	270.3	276.7	284	291.1	301.3	307.6
21	258.5	261.5	269.4	270.5	281.5	281.8	286.5	291.8	304.3
22	254.6	267.8	266.1	274.2	281.7	285.3	292.2	292.5	298.9
23	259.5	266.6	267.8	273.9	279.4	289.2	293.7	287.3	299.5
24	259.1	259.9	267.9	274.6	280.7	285.9	285	295.1	301.7
25	259.2	260	263.3	275.5	283.9	280.2	286.7	305.1	305.1
26	262.3	263.9	272	275.6	277.7	293.9	291.1	297	301.1
27	258.7	264.3	268.6	273	279.3	283.9	289	299.2	307.2
28	253.5	260.5	266.2	278	279.5	286.9	295.3	296.5	303.3
29	252.2	261.8	265.1	271.3	277.9	285	290.3	292.6	308.5
30	262.4	265.5	272.1	267.6	280.6	285.4	289.1	296.5	303.9

### A.3. Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik tanpa *Stemming*

**Tabel 0.3 Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik tanpa *Stemming***

Iterasi	Jumlah Topik								
	2	3	4	5	6	7	8	9	10
1	308.5	287.4	276.1	287.6	305.4	316.1	335.9	324.7	333.9
2	308.5	285.5	280.2	284.7	310.2	325.7	329.5	329.1	334.1
3	306.4	285.9	272.1	300.9	313.1	330.4	331.2	327.5	336.5
4	306.4	286.9	283.9	285.8	309.2	318.7	332.7	329.6	331.6
5	308.5	286.8	275.8	281.7	311	321.1	324	337.5	330.8
6	306.4	287.7	280.4	285.2	310.8	324.4	333.5	328.3	338.8
7	306.3	285.8	274.6	292.8	306.8	323.6	334.6	330.4	339.9
8	306.3	287.5	275.7	289	320.1	315.5	326.1	324.6	335.9

9	308.5	285	282.8	294.9	316.5	315.1	324.8	330	339.8
10	308.5	285.5	277.7	278.3	308.1	322.4	325.9	327.9	336.5
11	306.4	287.5	277.8	286.2	302.4	325.8	324.8	334.3	344
12	308.6	292.5	277.3	295.7	307.6	316.2	326	334.7	341.5
13	306.4	287.3	281.9	291.6	306.9	318.5	329.7	336.4	332.2
14	306.5	285.9	283.5	283	307.5	318.3	329.3	328.5	336.2
15	306.4	287.8	278.4	281	308.3	320.9	334.7	326.5	336.9
16	308	290.4	271.9	292.2	310.4	322.4	329.9	334.9	332.7
17	306.4	293.9	280.3	287.2	304.3	320.1	336.7	332.3	334.8
18	306.3	289	276.2	278.8	305.3	315.7	337.5	324.4	340.4
19	306.4	285.5	280.6	282.1	309.6	320.5	326.8	330.7	337.2
20	308.3	289.4	275	284.9	307.6	329.8	334.5	332.6	333.2
21	306.4	286.1	284	283.1	318.1	321.5	322.7	329.8	334.9
22	307.9	287.6	275.7	288.2	306	318.9	330.5	326.7	335.8
23	308.5	288.3	278.6	286.8	310.4	323.2	330.2	332.8	331.8
24	306.5	287.6	270.8	286.1	306.6	321.1	335	334	343.2
25	306.4	287.3	280.3	293	304.9	323.5	330.1	333	336.5
26	307.9	286.2	280.4	286.4	313.1	323.1	330.8	333.9	330.9
27	307.9	287.9	275.9	295.4	309.1	316.1	331.4	328.6	342.8
28	307.9	285.7	281.9	289.5	310.3	326.1	329.6	328	334.6
29	306.4	286.5	276.1	295.4	310.5	323.7	330.6	333.1	332.9
30	306.3	286.7	271.7	297.4	303.9	322.8	334.7	332.6	334

**Tabel 0.4 Hasil eksperimen analisis nilai perplexity untuk penentuan jumlah topik tanpa Stemming (lanjutan)**

Iterasi	Jumlah Topik								
	14	18	22	26	30	34	38	42	46
1	344	345.8	354.6	360.4	360.8	369.6	382.4	392	409
2	339.5	350.8	356.6	370.2	365.6	376.6	383.6	390.2	404.5

3	341.7	357.2	350.4	364.4	366	379	378	392.1	403.1
4	333	348.3	360	368.7	365.5	385.4	386.7	388.6	401.7
5	339.2	354.2	358	364.5	369.3	370.7	385.9	396.9	398.1
6	344.9	348.3	357.9	359.5	366.3	373.3	379.4	389.5	401.2
7	338.8	351.8	353.1	362.2	362.3	380.9	373.2	383.7	396
8	339.2	358.9	352.8	365.7	369.1	374.7	386.6	389.1	398
9	343.5	351.7	352.3	364.9	369.4	373.2	385.2	386.1	392.4
10	340.4	356.6	358	358.8	362.8	373.6	378	390.7	391.1
11	341.4	352.4	364.2	361.1	377.3	367.3	389.2	390.4	400.7
12	334.5	349.5	352.6	367.4	364.6	370.3	388.5	391.8	395.3
13	333.9	349.3	348.4	360	374.1	377.3	380.9	391.7	397.1
14	336.6	344.8	361.8	364.7	368.8	370	379.5	391.1	396.1
15	335.7	352.8	345.9	361.9	372.9	380	382.7	388.8	395.4
16	339.9	355.7	355.2	357.4	372.1	372.5	381.3	391.9	394.7
17	343.1	348.2	351.8	364.6	366	365.3	383.4	391.9	391.5
18	337.4	349.3	354.7	361.9	360.3	371.4	384.2	391.8	401
19	337.6	344	353.4	367.7	367.3	382	378.2	383.1	402.6
20	342.1	349.3	361	360.9	367	378.4	377	394.2	397.6
21	342.5	352.1	358.9	358.1	370.2	370	386.6	391.2	401.9
22	338.6	352.1	353.1	362.5	368.6	380.1	390.9	395.5	392.7
23	334.3	357.6	363.4	352.5	365.8	374.3	377.8	387.7	399.2
24	342.7	353.7	351.6	357.6	367	374.8	379.2	392.5	402.1
25	341.7	349	355.7	360.7	368.4	377.5	384.2	388.1	402.6
26	337.8	354	355.8	362.8	374.1	369.7	382	385.4	401.4
27	342.4	352.1	353.5	367.3	369.9	379.6	383.5	388.1	403.5
28	342.8	346	359.7	357.3	368.2	378.2	393.2	385.9	403.5
29	346.2	358.2	355	356.9	364.9	380.7	377.1	386.6	397.3
30	342.9	351.4	354.7	360.4	369.9	371.9	385.6	377.4	405.4

#### A.4. Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik 1

Soal	Jenis Uji Koherensi	Topik	Stem?	Skor	Skor (%)
1	<i>Word Intrusion Task</i>	0	Tidak	85	83.33%
2	<i>Word Intrusion Task</i>	0	Tidak	87	85.29%
3	<i>Word Intrusion Task</i>	0	Tidak	90	88.24%
4	<i>Word Intrusion Task</i>	0	Tidak	94	92.16%
5	<i>Word Intrusion Task</i>	0	Tidak	70	68.63%
6	<i>Word Intrusion Task</i>	1	Tidak	90	88.24%
7	<i>Word Intrusion Task</i>	1	Tidak	79	77.45%
8	<i>Word Intrusion Task</i>	1	Tidak	84	82.35%
9	<i>Word Intrusion Task</i>	1	Tidak	102	100.00%
10	<i>Word Intrusion Task</i>	1	Tidak	97	95.10%
11	<i>Word Intrusion Task</i>	2	Tidak	99	97.06%
12	<i>Word Intrusion Task</i>	2	Tidak	59	57.84%
13	<i>Word Intrusion Task</i>	2	Tidak	83	81.37%
14	<i>Word Intrusion Task</i>	2	Tidak	78	76.47%
15	<i>Word Intrusion Task</i>	2	Tidak	72	70.59%
16	<i>Word Intrusion Task</i>	3	Tidak	82	80.39%
17	<i>Word Intrusion Task</i>	3	Tidak	22	21.57%
18	<i>Word Intrusion Task</i>	3	Tidak	81	79.41%
19	<i>Word Intrusion Task</i>	3	Tidak	97	95.10%
20	<i>Word Intrusion Task</i>	3	Tidak	46	45.10%
21	<i>Word Intrusion Task</i>	0	Ya	90	88.24%
22	<i>Word Intrusion Task</i>	0	Ya	78	76.47%
23	<i>Word Intrusion Task</i>	0	Ya	78	76.47%
24	<i>Word Intrusion Task</i>	0	Ya	91	89.22%
25	<i>Word Intrusion Task</i>	0	Ya	57	55.88%
26	<i>Word Intrusion Task</i>	1	Ya	92	90.20%
27	<i>Word Intrusion Task</i>	1	Ya	84	82.35%
28	<i>Word Intrusion Task</i>	1	Ya	89	87.25%
29	<i>Word Intrusion Task</i>	1	Ya	94	92.16%
30	<i>Word Intrusion Task</i>	1	Ya	93	91.18%
31	<i>Word Intrusion Task</i>	2	Ya	87	85.29%
32	<i>Word Intrusion Task</i>	2	Ya	91	89.22%



33	<i>Word Intrusion Task</i>	2	Ya	86	84.31%
34	<i>Word Intrusion Task</i>	2	Ya	81	79.41%
35	<i>Word Intrusion Task</i>	2	Ya	76	74.51%
36	<i>Word Intrusion Task</i>	3	Ya	56	54.90%
37	<i>Word Intrusion Task</i>	3	Ya	89	87.25%
38	<i>Word Intrusion Task</i>	3	Ya	78	76.47%
39	<i>Word Intrusion Task</i>	3	Ya	80	78.43%
40	<i>Word Intrusion Task</i>	3	Ya	93	91.18%
41	<i>Topic Intrusion Task</i>	0	Tidak	75	73.53%
42	<i>Topic Intrusion Task</i>	1	Tidak	47	46.08%
43	<i>Topic Intrusion Task</i>	3	Tidak	43	42.16%
44	<i>Topic Intrusion Task</i>	0	Tidak	66	64.71%
45	<i>Topic Intrusion Task</i>	2	Tidak	67	65.69%
46	<i>Topic Intrusion Task</i>	2	Tidak	60	58.82%
47	<i>Topic Intrusion Task</i>	1	Tidak	83	81.37%
48	<i>Topic Intrusion Task</i>	3	Tidak	71	69.61%
49	<i>Topic Intrusion Task</i>	1	Tidak	76	74.51%
50	<i>Topic Intrusion Task</i>	2	Tidak	71	69.61%
51	<i>Topic Intrusion Task</i>	0	Tidak	67	65.69%
52	<i>Topic Intrusion Task</i>	3	Tidak	23	22.55%
53	<i>Topic Intrusion Task</i>	2	Tidak	68	66.67%
54	<i>Topic Intrusion Task</i>	3	Tidak	76	74.51%
55	<i>Topic Intrusion Task</i>	1	Tidak	71	69.61%
56	<i>Topic Intrusion Task</i>	0	Tidak	69	67.65%
57	<i>Topic Intrusion Task</i>	3	Tidak	21	20.59%
58	<i>Topic Intrusion Task</i>	0	Tidak	67	65.69%
59	<i>Topic Intrusion Task</i>	1	Tidak	61	59.80%
60	<i>Topic Intrusion Task</i>	2	Tidak	71	69.61%
61	<i>Topic Intrusion Task</i>	3	Ya	58	56.86%
62	<i>Topic Intrusion Task</i>	2	Ya	74	72.55%
63	<i>Topic Intrusion Task</i>	2	Ya	76	74.51%
64	<i>Topic Intrusion Task</i>	3	Ya	68	66.67%
65	<i>Topic Intrusion Task</i>	1	Ya	77	75.49%
66	<i>Topic Intrusion Task</i>	0	Ya	71	69.61%
67	<i>Topic Intrusion Task</i>	0	Ya	75	73.53%

68	<i>Topic Intrusion Task</i>	3	Ya	66	64.71%
69	<i>Topic Intrusion Task</i>	1	Ya	69	67.65%
70	<i>Topic Intrusion Task</i>	1	Ya	71	69.61%
71	<i>Topic Intrusion Task</i>	2	Ya	77	75.49%
72	<i>Topic Intrusion Task</i>	3	Ya	68	66.67%
73	<i>Topic Intrusion Task</i>	0	Ya	71	69.61%
74	<i>Topic Intrusion Task</i>	3	Ya	70	68.63%
75	<i>Topic Intrusion Task</i>	2	Ya	56	54.90%
76	<i>Topic Intrusion Task</i>	1	Ya	43	42.16%
77	<i>Topic Intrusion Task</i>	0	Ya	69	67.65%
78	<i>Topic Intrusion Task</i>	2	Ya	47	46.08%
79	<i>Topic Intrusion Task</i>	0	Ya	75	73.53%
80	<i>Topic Intrusion Task</i>	1	Ya	72	70.59%

### A.5. Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik 2

Soal	Uji	Topik	Stem?	Skor	Skor (%)
41	<i>Topic Intrusion Task</i>	0	Tidak	80	65.57
42	<i>Topic Intrusion Task</i>	1	Tidak	19	15.57
43	<i>Topic Intrusion Task</i>	3	Tidak	42	34.43
44	<i>Topic Intrusion Task</i>	0	Tidak	89	72.95
45	<i>Topic Intrusion Task</i>	2	Tidak	85	69.67
46	<i>Topic Intrusion Task</i>	2	Tidak	92	75.41
47	<i>Topic Intrusion Task</i>	1	Tidak	95	77.87
48	<i>Topic Intrusion Task</i>	3	Tidak	82	67.21
49	<i>Topic Intrusion Task</i>	1	Tidak	99	81.15
50	<i>Topic Intrusion Task</i>	2	Tidak	79	64.75
51	<i>Topic Intrusion Task</i>	0	Tidak	84	68.85
52	<i>Topic Intrusion Task</i>	3	Tidak	24	19.67
53	<i>Topic Intrusion Task</i>	2	Tidak	94	77.05
54	<i>Topic Intrusion Task</i>	3	Tidak	85	69.67
55	<i>Topic Intrusion Task</i>	1	Tidak	88	72.13
56	<i>Topic Intrusion Task</i>	0	Tidak	75	61.48
57	<i>Topic Intrusion Task</i>	3	Tidak	24	19.67
58	<i>Topic Intrusion Task</i>	0	Tidak	95	77.87
59	<i>Topic Intrusion Task</i>	1	Tidak	63	51.64
60	<i>Topic Intrusion Task</i>	2	Tidak	86	70.49
61	<i>Topic Intrusion Task</i>	3	Ya	94	77.05
62	<i>Topic Intrusion Task</i>	2	Ya	85	69.67
63	<i>Topic Intrusion Task</i>	2	Ya	90	73.77
64	<i>Topic Intrusion Task</i>	3	Ya	86	70.49
65	<i>Topic Intrusion Task</i>	1	Ya	93	76.23
66	<i>Topic Intrusion Task</i>	0	Ya	76	62.30
67	<i>Topic Intrusion Task</i>	0	Ya	99	81.15
68	<i>Topic Intrusion Task</i>	3	Ya	89	72.95
69	<i>Topic Intrusion Task</i>	1	Ya	84	68.85
70	<i>Topic Intrusion Task</i>	1	Ya	92	75.41
71	<i>Topic Intrusion Task</i>	2	Ya	69	56.56

72	<i>Topic Intrusion Task</i>	3	Ya	78	63.93
73	<i>Topic Intrusion Task</i>	0	Ya	72	59.02
74	<i>Topic Intrusion Task</i>	3	Ya	73	59.84
75	<i>Topic Intrusion Task</i>	2	Ya	94	77.05
76	<i>Topic Intrusion Task</i>	1	Ya	27	22.13
77	<i>Topic Intrusion Task</i>	0	Ya	92	75.41
78	<i>Topic Intrusion Task</i>	2	Ya	68	55.74
79	<i>Topic Intrusion Task</i>	0	Ya	22	18.03
80	<i>Topic Intrusion Task</i>	1	Ya	88	72.13

### A.5. Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden

**Tabel 0.5 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden Word Intrusion Task**

Responden	Skor T0	Skor T1	Skor T2	Skor T3	Total Skor (#)	Total Skor (%)
1	10	10	10	10	40	100
2	7	8	9	6	30	75
3	9	10	10	7	36	90
4	10	10	9	7	36	90
5	10	9	9	9	37	92.5
6	10	10	10	7	37	92.5
7	6	6	6	4	22	55
8	10	10	7	9	36	90
9	9	10	9	9	37	92.5
10	4	3	3	5	15	37.5
11	9	10	9	8	36	90
12	10	10	8	6	34	85
13	7	9	8	8	32	80
14	6	10	10	9	35	87.5

15	8	9	8	6	31	77.5
16	7	9	8	8	32	80
17	10	10	10	8	38	95
18	9	10	10	7	36	90
19	7	9	5	3	24	60
20	9	10	9	6	34	85
21	9	10	10	8	37	92.5
22	9	10	8	9	36	90
23	7	9	9	6	31	77.5
24	9	9	9	5	32	80
25	7	9	9	7	32	80
26	9	8	6	6	29	72.5
27	9	10	9	9	37	92.5
28	9	9	8	6	32	80
29	9	10	6	9	34	85
30	3	8	9	7	27	67.5
31	7	9	9	9	34	85
32	6	6	4	3	19	47.5
33	9	9	8	7	33	82.5
34	3	10	8	7	28	70
35	8	10	10	9	37	92.5
36	9	9	7	9	34	85
37	8	2	5	3	18	45
38	5	9	8	7	29	72.5
39	10	8	8	6	32	80
40	8	8	8	8	32	80
41	6	9	8	8	31	77.5
42	10	10	9	7	36	90
43	6	10	9	8	33	82.5
44	7	7	4	4	22	55
45	8	7	6	8	29	72.5
46	8	9	9	6	32	80
47	7	10	8	6	31	77.5

48	7	7	7	9	30	75
49	9	10	7	8	34	85
50	8	10	9	6	33	82.5
51	6	10	9	9	34	85
52	7	6	5	5	23	57.5
53	9	9	9	8	35	87.5
54	6	8	7	5	26	65
55	10	10	9	8	37	92.5
56	10	10	8	7	35	87.5
57	7	9	7	8	31	77.5
58	8	10	9	8	35	87.5
59	8	7	6	7	28	70
60	10	10	7	9	36	90
61	9	10	10	8	37	92.5
62	8	10	7	8	33	82.5
63	8	9	9	8	34	85
64	10	10	9	9	38	95
65	9	10	10	8	37	92.5
66	9	9	7	8	33	82.5
67	8	9	8	7	32	80
68	8	9	5	5	27	67.5
69	9	10	9	8	36	90
70	9	5	6	4	24	60
71	5	6	4	6	21	52.5
72	10	10	10	9	39	97.5
73	7	7	5	3	22	55
74	7	10	7	8	32	80
75	9	10	8	8	35	87.5
76	6	6	6	4	22	55
77	8	10	9	8	35	87.5
78	9	9	7	8	33	82.5
79	10	9	9	8	36	90
80	10	8	8	8	34	85

81	10	10	10	9	39	97.5
82	7	9	10	7	33	82.5
83	5	8	7	8	28	70
84	7	7	10	7	31	77.5
85	8	10	8	8	34	85
86	7	8	5	6	26	65
87	9	10	8	6	33	82.5
88	10	10	10	8	38	95
89	6	10	9	6	31	77.5
90	9	9	7	7	32	80
91	9	8	9	8	34	85
92	10	9	10	9	38	95
93	9	10	8	8	35	87.5
94	6	6	4	3	19	47.5
95	7	10	7	6	30	75
96	8	10	10	8	36	90
97	9	10	9	8	36	90
98	7	10	10	8	35	87.5
99	5	2	3	0	10	25
100	9	10	9	8	36	90
101	8	10	10	9	37	92.5
102	10	10	9	8	37	92.5

**Tabel 0.6 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik Berbasis Responden Topic Intrusion Task 1**

Responden	Skor T0	Skor T1	Skor T2	Skor T3	Total Skor (#)	Total Skor (%)
1	2	2	2	4	10	25
2	8	6	8	9	30	75
3	3	5	3	2	14	35

4	3	6	2	4	15	37.5
5	9	8	10	7	33	82.5
6	8	7	9	8	31	77.5
7	4	2	2	3	11	27.5
8	8	10	10	9	36	90
9	10	8	10	9	36	90
10	3	2	4	8	17	42.5
11	2	1	3	4	10	25
12	9	7	10	7	32	80
13	8	8	9	3	27	67.5
14	10	8	10	10	37	92.5
15	8	8	8	9	32	80
16	7	7	9	7	29	72.5
17	9	7	10	10	35	87.5
18	9	7	9	10	34	85
19	4	3	5	2	13	32.5
20	8	8	7	8	30	75
21	8	8	10	9	34	85
22	9	10	9	10	37	92.5
23	7	8	9	9	32	80
24	9	8	9	10	35	87.5
25	8	8	10	7	32	80
26	7	7	10	10	33	82.5
27	8	8	10	9	34	85
28	7	8	9	7	30	75
29	1	2	3	4	10	25
30	9	8	10	8	34	85
31	10	9	10	8	36	90
32	2	4	4	1	11	27.5
33	4	2	4	1	10	25
34	1	2	5	1	9	22.5
35	10	8	10	10	37	92.5
36	9	9	9	10	36	90



37	4	3	3	3	13	32.5
38	9	4	9	7	28	70
39	8	8	10	9	34	85
40	7	4	3	4	19	47.5
41	2	1	3	2	8	20
42	8	7	10	8	32	80
43	4	1	2	2	9	22.5
44	3	3	5	7	17	42.5
45	2	1	2	1	7	17.5
46	8	8	9	8	32	80
47	7	8	8	10	32	80
48	2	3	1	1	6	15
49	9	7	9	10	34	85
50	2	3	4	4	13	32.5
51	8	8	10	9	34	85
52	3	2	5	5	15	37.5
53	5	5	7	5	21	52.5
54	2	4	2	0	9	22.5
55	8	7	9	6	29	72.5
56	6	2	6	4	17	42.5
57	2	3	2	3	10	25
58	7	9	10	9	34	85
59	8	1	3	2	13	32.5
60	4	3	4	3	13	32.5
61	9	6	9	8	31	77.5
62	8	8	10	8	33	82.5
63	2	6	2	3	12	30
64	10	7	9	7	32	80
65	8	7	9	7	30	75
66	9	6	10	10	34	85
67	3	1	1	2	7	17.5
68	8	2	7	7	23	57.5
69	3	3	2	1	10	25

70	3	1	0	2	6	15
71	0	4	1	4	9	22.5
72	9	8	9	7	32	80
73	6	9	4	7	25	62.5
74	5	6	9	8	27	67.5
75	7	7	8	9	30	75
76	8	8	8	8	31	77.5
77	5	8	7	3	22	55
78	2	2	1	5	10	25
79	7	4	10	6	26	65
80	6	8	10	8	31	77.5
81	10	8	10	10	37	92.5
82	10	7	10	10	36	90
83	7	7	7	6	26	65
84	7	7	9	8	30	75
85	8	9	10	6	32	80
86	6	6	6	7	24	60
87	10	6	3	3	21	52.5
88	8	8	10	10	35	87.5
89	9	3	10	9	30	75
90	8	8	10	8	33	82.5
91	8	10	9	9	35	87.5
92	6	8	9	8	30	75
93	10	10	10	8	37	92.5
94	4	1	2	0	8	20
95	3	2	3	3	10	25
96	9	8	9	7	32	80
97	9	6	7	9	30	75
98	9	8	9	7	32	80
99	6	3	1	4	14	35
100	9	9	8	10	36	90
101	8	8	10	7	32	80
102	8	7	10	10	34	85

**Tabel 0.7 Rekapitulasi Tanggapan Kuesioner Uji Koherensi Topik  
Berdasarkan Responden Topic Intrusion Task 2**

Responden	Skor T0	Skor T1	Skor T2	Skor T3	Total Skor (#)	Total Skor (%)
1	7	6	9	6	27	67.5
2	7	7	9	4	26	65
3	7	7	9	4	26	65
4	8	8	10	9	34	85
5	8	8	10	6	31	77.5
6	9	8	9	5	30	75
7	8	8	9	3	27	67.5
8	2	2	7	4	14	35
9	4	2	3	4	13	32.5
10	8	7	6	8	28	70
11	3	5	6	4	17	42.5
12	9	10	10	8	36	90
13	5	2	3	3	13	32.5
14	3	2	4	0	10	25
15	5	4	9	4	21	52.5
16	8	7	10	7	31	77.5
17	8	8	10	7	32	80
18	8	9	8	7	32	80
19	9	5	9	8	30	75
20	8	6	10	7	30	75
21	8	5	9	3	24	60
22	8	5	10	10	32	80
23	8	8	9	9	33	82.5
24	8	5	10	6	28	70
25	1	0	0	4	6	15
26	2	2	3	3	10	25

27	4	3	7	7	20	50
28	6	7	9	7	28	70
29	5	3	1	1	11	27.5
30	8	8	10	7	32	80
31	6	9	10	5	29	72.5
32	9	8	10	7	33	82.5
33	0	3	2	1	6	15
34	3	3	1	3	10	25
35	3	3	1	3	10	25
36	7	7	8	6	27	67.5
37	1	2	0	3	6	15
38	7	7	8	5	27	67.5
39	9	6	9	7	30	75
40	6	5	6	6	23	57.5
41	9	8	10	8	34	85
42	7	5	7	6	24	60
43	8	7	10	6	30	75
44	8	5	10	6	28	70
45	1	5	1	2	9	22.5
46	2	5	1	1	10	25
47	8	5	8	7	27	67.5
48	1	1	3	5	10	25
49	7	8	8	3	25	62.5
50	2	1	1	2	6	15
51	8	8	10	7	32	80
52	8	8	10	6	31	77.5
53	2	3	5	3	12	30
54	8	8	5	3	23	57.5
55	5	6	8	5	23	57.5
56	8	7	7	6	27	67.5
57	8	6	9	8	30	75
58	2	1	5	3	11	27.5
59	8	7	7	7	28	70

60	7	8	10	8	32	80
61	5	3	4	4	16	40
62	5	5	10	8	27	67.5
63	5	3	2	3	14	35
64	5	7	9	3	23	57.5
65	7	3	7	8	24	60
66	3	2	4	3	12	30
67	7	6	8	5	25	62.5
68	7	7	7	7	27	67.5
69	9	7	10	9	34	85
70	5	5	6	6	21	52.5
71	7	9	9	8	32	80
72	6	5	5	4	19	47.5
73	3	4	3	5	14	35
74	0	2	2	3	7	17.5
75	9	9	10	5	32	80
76	7	7	10	8	31	77.5
77	8	7	10	8	32	80
78	9	7	10	6	31	77.5
79	5	7	9	7	27	67.5
80	8	8	9	6	30	75
81	6	8	10	7	30	75
82	8	10	10	8	35	87.5
83	6	7	10	5	27	67.5
84	7	8	9	6	29	72.5
85	7	8	10	7	31	77.5
86	6	7	7	7	26	65
87	8	7	9	9	32	80
88	3	2	2	7	13	32.5
89	2	2	2	3	9	22.5
90	9	8	9	6	31	77.5
91	8	7	9	8	31	77.5
92	8	6	10	6	29	72.5

93	7	5	9	7	27	67.5
94	3	4	2	4	14	35
95	8	7	6	6	26	65
96	9	10	10	8	36	90
97	8	8	10	5	30	75
98	7	7	10	8	31	77.5
99	7	8	9	8	31	77.5
100	8	7	9	7	30	75
101	8	7	8	5	27	67.5
102	8	5	5	4	21	52.5
103	6	10	8	8	31	77.5
104	6	6	8	6	25	62.5
105	3	1	3	3	10	25
106	7	7	10	7	30	75
107	9	7	10	9	34	85
108	7	8	7	4	26	65
109	6	6	8	3	22	55
110	6	2	7	5	19	47.5
111	8	7	8	7	30	75
112	8	9	6	7	30	75
113	7	6	10	6	28	70
114	5	4	3	5	17	42.5
115	3	1	2	2	8	20
116	7	7	7	5	25	62.5
117	7	8	9	5	28	70
118	7	8	9	7	31	77.5
119	6	8	9	7	29	72.5
120	7	7	8	2	24	60
121	9	6	10	8	32	80
122	7	7	9	7	29	72.5

## **LAMPIRAN B**